VOLUME 23 OCTOBER, 1935 NUMBER 10

PROCEEDINGS
of
The Institute of Radio Engineers

Application Blank for Associate Membership on Page XVI

Institute of Radio Engineers Forthcoming Meetings

CINCINNATI SECTION
October 15, 1935

DETROIT SECTION
October 18, 1935

LOS ANGELES SECTION
October 15, 1935

NEW YORK MEETING
October 23, 1935

PHILADELPHIA SECTION
October 3, 1935

WASHINGTON SECTION
October 14, 1935

Radio Emissions of Standard Frequency

The National Bureau of Standards provides standard frequency emissions from its station WWV at Beltsville, Md. Beginning October 1, 1935, on each Tuesday and Friday the emissions are continuous unmodulated waves and on each Wednesday they are modulated by an audio frequency, generally 1000 cycles. There are no emissions on legal holidays.

On all schedules three radio carrier frequencies are transmitted as follows: noon to 1 P.M., Eastern Standard Time, 15,000 kilocycles; 1:15 to 2:15 P.M., 10,000 kilocycles; and 2:30 to 3:30 P.M., 5000 kilocycles. The accuracy of these frequencies will at all times be better than a part in five million.

During the first five minutes of each transmission announcements are given of the station call letters, the frequency of transmission, and the frequency of modulation, if any. For the CW emissions, the announcements are in telegraphic code and are repeated at ten-minute intervals. For the modulated emissions, the announcements are given by voice only at the beginning of each carrier frequency transmission, the remainder of the hour being an uninterrupted audio frequency. The CW emissions are from a twenty-kilowatt transmitter and the modulated transmissions are from a one-kilowatt set.

Information on how to utilize these signals is given in a pamphlet obtainable on request from the National Bureau of Standards, Washington, D. C. Reports from those using this service will be welcomed by the Bureau. As the modulated emissions are somewhat experimental it is particularly desired that users report their experiences outlining methods of utilization, information on relative fading, intensity, etc., on the three carrier frequencies and preferences as to the audio frequency to be furnished.

Committee Work

Admissions Committee

The Admissions Committee met on September 10 in the Institute office and those present were Austin Bailey, chairman, R. A. Heising, L. C. F. Horle, and H. P. Westman, secretary. Three applications for admission and two for transfer to the Fellow grade were approved. Of

six applications for admission to the grade of Member, three were approved, two were tabled pending additional information, and one was denied. Nine applications for transfer to the grade of Member were considered Six of these were approved and the remaining three were denied.

Sections Committee

The annual meeting of the Sections Committee was held in the Hotel Statler in Detroit on July 1. This was during the Tenth Annual Convention and those present were C. J. Banfer, A. B. Buchanan, H. L. Byerlay, E. D. Cook, Alfred Crossley, E. C. Denstaedt, William Dent, G. E. Frater, H. S. Gould, V. M. Graham, R. A. Hackbusch, George Lewis, Knox McIlwain, B. B. Minnium, L. M. Price, F. H. Rousch, F. E. Terman, Irving Wolff, and H. P. Westman, secretary.

An analysis of the membership and the number of meetings held by each section was made and considered to indicate satisfactory conditions. Financial reports were reviewed and because several sections did not submit these reports it was recommended that the Board of Directors authorize the secretary to withhold future rebates on both members and meetings to those sections which do not submit financial reports when they are due.

A general discussion was held on the subjects of obtaining applications for transfer to higher grades from members who are qualified for increases in membership rating and possible sources of new members.

STANDARDIZATION

Technical Committee on Radio Receivers—IRE

A meeting of the Technical Committee on Radio Receivers was held on July 11 in the Institute office and those present were H. A. Wheeler, chairman, C. B. Aiken, L. F. Curtis, E. T. Dickey, J. F. Dreyer, M. T. Smith (representing A. E. Thiessen), and H. P. Westman, secretary.

This was the first meeting of the committee and its organization and scope of operation were considered in detail. It was agreed that several subcommittees be appointed to work on the various portions of the existing report. In anticipation of this, the report was reviewed in a general way for the benefit of the subcommittees.

Technical Committee on Transmitters and Antennas—IRE

The Subcommittee on Transmitters operating under the Technical Committee on Transmitters and Antennas of the Institute met at the Institute office on August 9. Those present were D. G. Little, chair-

man, H. A. Chinn (representing A. B. Chamberlain), E. B. Ferrell, E. G. Ports, D. S. Rau, and H. P. Westman, secretary.

At this meeting the committee reviewed all of the existing material on transmitters and prepared a final report for submission to the Technical Committee on Transmitters and Antennas.

Technical Committee on Electronics—IRE

The Subcommittee on Electron Beam and Miscellaneous Tubes of the Technical Committee on Electronics met at the RCA Radiatron plant in Harrison, N. J., on September 9. Those present were G. F. Metcalf, chairman, H. N. Blackmon (representing Lee Sutherlin), A. B. DuMont, M. S. Glass, Ben Kievit, Jr., T. B. Perkins, B. J. Thompson, and H. P. Westman, secretary.

The committee devoted its time to the preparation of definitions applying to cathode-ray tubes.

Institute Meetings

Boston Section

The annual meeting of the Boston Section was held on May 24 at Massachusetts Institute of Technology and the meeting was presided over by E. L. Chaffee, chairman. Fifteen were present at the informal dinner which preceded the meeting and 100 were in attendance at it.

Two papers were presented at this meeting, the first being on "An Investigation of the Radiation of Sound from an Intense Source" by W. M. Hall, an instructor in the Communication Division of the Electrical Engineering Department of Massachusetts Institute of Technology. The second paper was "An Investigation of the Carbon Microphone" by D. N. Truscott, Commonwealth Fund Fellow of Massachusetts Institute of Technology.

In the election of officers, E. L. Bowles, Professor of Electrical Engineering at Massachusetts Institute of Technology was designated chairman; H. R. Mimno of Harvard University was named vice chairman; and R. G. Porter, Associate Professor of Electrical Fngineering of Northeastern University was re-elected secretary-treasurer.

Detroit Section

The May meeting of the Detroit Section was held on the 24th with A. B. Buchanan, chairman, presiding. It was held in the Detroit News Conference Room and was attended by forty-five members and guests.

A paper on "Experiments in Nuclear Physics" was presented by E. R. Gaerttner, graduate research student of the physics department, University of Michigan.

He pointed out that the study of the atomic nucleus is particularly interesting because the energy of the atom is contained in the nucleus. The disintegration of an atom results from bombarding it with hydrogen ions or protons, heavy hydrogen ions or deutrons, and doubly ionized helium ions or alpha particles. These are accelerated by passing through a field of about a million volts to produce greater ionic currents than can be obtained from radioactive materials. Neutrons which are of nearly the same mass as hydrogen but have no electrical charge and gamma rays which are very hard X rays may also be used for disintegrating atoms.

Several methods were outlined for obtaining high direct voltages to impart to the ions the necessary velocity to break up the atom. The particles given off as a result of this bombardment can be made to effect electrical circuits and records made of the rate at which they are emitted.

The June meeting of the section, also held in the Detroit News Conference Room and presided over by Chairman Buchanan, was on the 21st. Sixty-five were present and twenty attended the informal dinner which preceded it.

E. C. Denstaedt, Supervisor of Radio of the Detroit Police Department presented a paper on "Progress in Ultra-High-Frequency Communication Equipment." In it he reviewed present police equipment and discussed the need for two-way communication apparatus by both police and fire departments. The older two-way equipment using modulated oscillators for transmission and superheterodyne receivers was described and its limitations shown. Some recently developed equipment was described and demonstrated and included a seven and one-half-meter superheterodyne receiver as well as a superregenerative and a tuned radio-frequency receiver for the same frequency range. These receivers used the acorn type pentode tubes and were decidedly better than the average superregenerative receiver when used at a fixed station. They were not used in automobiles.

A number of those present participated in the discussion and the meeting was terminated with an inspection of various transmitters and receivers which were on display and a demonstration of a portable car transmitter.

San Francisco Section

The San Francisco Section met at the Bellevue Hotel. The meeting was presided over by Robert Kirkland, vice chairman, and attended by fifty-four. Twenty were present at the informal dinner which preceded the meeting.

A paper on "Performance, Tests, and Trends of Loud Speaker Design" was presented by H. S. Knowles, Chief Engineer of the Jensen Radio Manufacturing Company. He discussed first some of the fundamentals of speech and sound and their effect on loud speaker design. He then described methods used to test loud speaker performance and analyzed typical response curves pointing out reasons for the variations in response with frequency. The polar characteristics were mentioned. Recent trends in loud speaker design especially in regard to high fidelity reproduction of speech and music were then covered.

Ralph Sherman, chairman of the section, who has recently changed his business address found it necessary to resign and a new slate of officers was chosen. Robert Kirkland of the Mackay Radio and Telegraph Company, formerly vice chairman, was elected chairman; V. J. Freiermuth of the Pacific Telephone and Telegraph Company, formerly secretary-treasurer, was named vice chairman; and Henry Tanck of RCA Communications, Inc., was elected secretary-treasurer.

Personal Mention

C. R. Banks has become a test engineer for the RCA Manufacturing Company, Camden, having formerly been connected with the De Forest Radio Company.

W. J. Cahill, previously with Radio Receptor Company, is now at the Naval Research Laboratory, Anacostia, D. C.

O. H. Caldwell is now editor of a new publication "*Radio Today*" in New York City, having formerly been associated with the McGraw-Hill Publishing Company.

Formerly with H. H. Horn Radio Manufacturing Company, D. D. Dressen is now a geophysical engineer for the Rieber Laboratory of Los Angeles, Calif.

W. B. Goulett, Lieutenant, U. S. N., has been transferred from the U. S. S. *Arizona* to the Bureau of Engineering in Washington, D. C.

C. F. Holden, Lieutenant, U. S. N., has been transferred from the U. S. S. *Tarbell* to the Naval Radio Station, Wailupe, Oahu, T. H.

C. J. King, Jr., Lieutenant, U. S. A., has been transferred from Fort Monmouth to West Point.

H. M. McClelland, Captain, U. S. Air Corps, has been transferred from Coronado, Calif., to Montgomery, Ala.

R. G. H. Meyer, Lieutenant, U. S. A., is now at Fort Lewis, Wash., having previously been at Fort Sam Houston, Texas.

G. M. Neely, Lieutenant, U. S. N., has been transferred from Washington, D. C. to the U. S. S. *Roper* basing at San Diego, Calif.

Previously with Lear Developments, Inc., E. H. Potter has joined the radio engineering staff of the RCA Manufacturing Company of Camden, N. J.

L. F. Pries is now with the Wurlitzer Grand Piano Company of De Kalb, Ill., having formerly been employed by the Colonial Radio Shop.

C. W. Richard, civilian radio engineer for the U. S. Air Corps, has been transferred from Fort Monmouth to Chanute Field, Rantoul, Ill.

F. X. Rettenmeyer has left Bell Telephone Laboratories, Inc., and joined the engineeting staff of RCA Manufacturing Company, Camden, N. J.

L. C. Smeby has left KSTP and has become technical supervisor of WXYZ in Detroit.

Previously with All-American Cables, H. O. Storm has joined the engineering staff of Globe Wireless Communications in San Francisco.

L. E. Swedlund, formerly with Westinghouse Electric and Manufacturing Company, has joined the research and development laboratory of the RCA Manufacturing Company, Harrison, N. J.

R. A. Wood of RCA Communications has been transferred from Point Reyes to Huntington Beach, Calif.

A. C. Wooldridge has left *American Radio News* to become a transmitter design engineer for Lear Developments.

R. A. Woolverton, Captain, U. S. A., has been transferred from San Francisco, Calif., to Omaha, Neb.

Proceedings of the Institute of Radio Engineers
Volume 23, Number 10 October, 1935

TECHNICAL PAPERS

AUTOMATIC FREQUENCY CONTROL*

By

Charles Travis

(Formerly RCA License Laboratory, New York City; now, Philco Radio and Television Corporation, Philadelphia, Pennsylvania)

Summary—*A system is described for electronic control of the local oscillator frequency in a superheterodyne receiver for the purpose of centering the signal carrier in the intermediate-frequency band in spite of inaccuracies of manual tuning and oscillator drift. The system consists of two parts, a discriminator for converting frequency departures into direct voltage differences, and a control circuit for converting voltage differences into reactance variations of the oscillator circuit. The system thus tends automatically to control the intermediate frequency produced, so that maladjustments which otherwise would produce serious alignment errors will result in actual error of only a few cycles. The system also makes practicable remote tuning control and other simple mechanical tuning methods.*

A HIGH degree of selectivity in a receiver is of no great value unless it is possible to tune the set with a corresponding degree of fineness or accuracy and thereafter to maintain this accuracy; indeed, if the tuning is very inadequate the high selectivity may be worse than useless. In the all-wave set, selectivity has been effectively increased fifteen to twenty times over that usual for broadcast reception, merely because the received frequencies have been increased by that amount without changing the intermediate-frequency band width; at twenty megacycles the nominal ten-kilocycle intermediate-frequency band is only 0.05 per cent of the base frequency. To meet this increase in selectivity in present-day receivers manual tuning means have been improved by the employment of more smoothly working speed-reducing movements to operate the variable condenser gang, but the maintenance of proper tuning, after the station signal has once been correctly brought in, is a problem that has as yet no satisfactory solution. Oscillator drift, if not corrected by more or less frequent manual readjustment, is capable of mistuning the signal by many channels in the course of a few hours' run.

In the broadcast band, conditions are quite as serious if quality of reproduction is a consideration, and the trend in high fidelity is bringing the same problem up in another form. It seems to be true that the aver-

* Decimal classification: R361.2. Original manuscript received by the Institute, May 10, 1935. Presented before Tenth Annual Convention, Detroit, Mich., July 1, 1935.

age listener does not tune his set well enough to obtain the best quality it is capable of giving, partly from negligence, and partly from lack of the necessary skill, in which case the mechanical design of the set is a possible contributing factor.

These considerations, among others, indicate a need for some way of supplementing the accuracy of manual tuning by more or less automatic means—some method of bringing the signal carrier precisely to the center of its intermediate-frequency band and anchoring it there in spite of small original maladjustments of tuning or others that subsequently arise from thermal changes and the like. This would be the function of an automatic frequency control system such as this paper describes.

In addition to the field of usefulness just pointed out, in connection with a manually tuned receiver, the automatic frequency control system would have a further application to sets with various forms of mechanical tuning—remote control or automatic station preselection. These tuning devices, unless very finely (and expensively) made, are likely to be inadequate from the standpoint of accuracy of tuning. Automatic frequency control would provide this accuracy after the mechanical device had roughly performed the selection of the signal.

An automatic frequency control system will consist of two distinct units; a frequency discriminator or frequency sensitive detector that generates a bias varying with changes of the intermediate-frequency signal carrier frequency, and a control unit that is acted upon by this bias, and whose function is to vary the local oscillator frequency accordingly. The two units are so coördinated that if the intermediate-frequency carrier tends to move away from the mid-band position, the oscillator frequency changes sufficiently to restore proper alignment. A close analogy to the operation of an automatic volume control system is to be noted.

At this point the system may develop along either of two radically different lines. The discriminating unit will necessarily be electronic (a tube or tubes), but the control unit may be an electronic device or alternatively a mechanical device, electrodynamic in nature. Prior art discloses means for varying oscillator frequency by the use of what is essentially a direct-current meter movement, deflection of which varies certain physical elements of the oscillating circuit. A relay of some kind will accomplish the same purpose.

While electronic methods have likewise been disclosed in prior art, it appears that the inefficiency of the proposed methods or inherent limitations, in one way or another, have rendered them impracticable for use in conjunction with the present wide band receiver. It is only

with the development of circuits capable of swinging the oscillator frequency by many parts per hundred that the electronic method has been put into a position to compete with the electrodynamic method.

With the two ways of accomplishing the purpose put upon an equal footing in regard to range and sensitivity, it is far too early to predict which one will meet with the more general acceptance. The not-too-optimistic engineer will foresee that either method will bring out difficulties; there are many ways in which an electrodynamic device can go wrong, and there are just about as many ways in which electron tubes can do likewise. It seems that only parallel development of the two will indicate which is better.

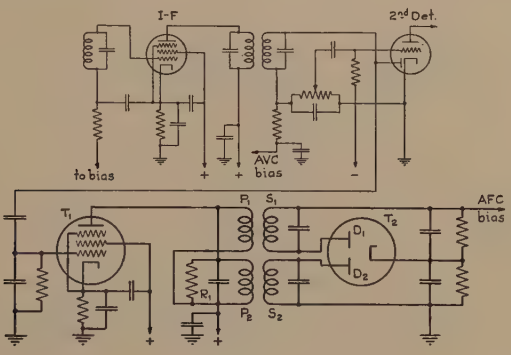


Fig. 1

The electronic method has the initial advantage that it adds no additional moving parts to the receiver. In rebuttal it can be pointed out that the modern receiver already has many mechanical movements, in the variable condenser gang and band changing switch, in the level and tone controls, and in some instances in an intermediate-frequency band width control. In fairness it must be pointed out, however, that these devices are not automotive, and their possible ruggedness is limited only by the strength of the prospective user.

The writer therefore holds no brief for one method against the other. The electronic method alone is discussed here because of the peculiar nature and interest of the problems arising; if parallel study of the mechanical method is to be made it likewise will be in order to describe this separately.

In the development of the system here described the form taken by the discriminator is fairly well crystallized. A differential rectifier is used, following closely a disclosure by Round.[1] A schematic diagram of the present circuit is shown in Fig. 1, with a variant in Fig. 2.

[1] U. S. Patent 1,642,173.

A single side rectifier with a single off-tune circuit in its input might be used, operation taking place upon the sloping side of a resonance curve, but such a device is open to the objection that the signal amplitude would affect the center value of the bias output. Accordingly the differential circuit shown in Figs. 1 or 2 is to be preferred.

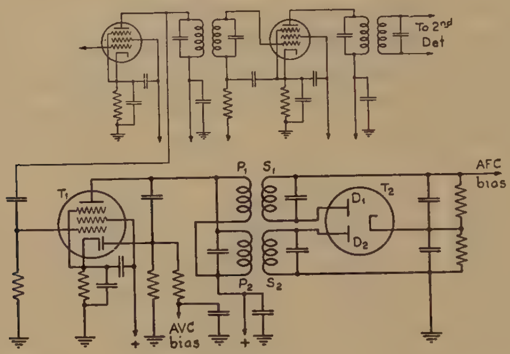


Fig. 2

Two similar intermediate-frequency transformers have their primaries P_1 and P_2 connected in parallel in the plate of a driver tube T_1 (type 78 or 6D6), and this composite primary is tuned to the mid-intermediate frequency. The secondaries S_1 and S_2, loosely coupled to the primaries, are tuned to different frequencies, lying above and below the mid-frequency by equal amounts. To help decouple the secondaries from each other and so to cause them to act like isolated single tuned circuits, the primary is damped by the shunt resistance R_1 of the order of 0.1 megohm. In Fig. 2 an automatic volume control diode is driven from the primary, and the diode load takes the place of R_1.

The secondaries S_1 and S_2 are each connected into one of the plates of a double diode T_2 (at present a type 85 with triode section unused). The cathode of this tube is "floating" for direct current; one diode return connects to ground or to a point of fixed potential determining the initial bias, and the automatic frequency control bias output is taken from the other diode return.

The output of this detector is the algebraic difference between the rectified outputs of the two diodes. If the intermediate-frequency carrier is off center frequency towards the resonance of S_1, diode D_1 will produce the greater output of the two and the generated automatic frequency control bias will be negative with respect to the initial value. Conversely if the carrier is off in the other direction the reverse will be

true. When the carrier is exactly aligned the bias will equal the initial value.

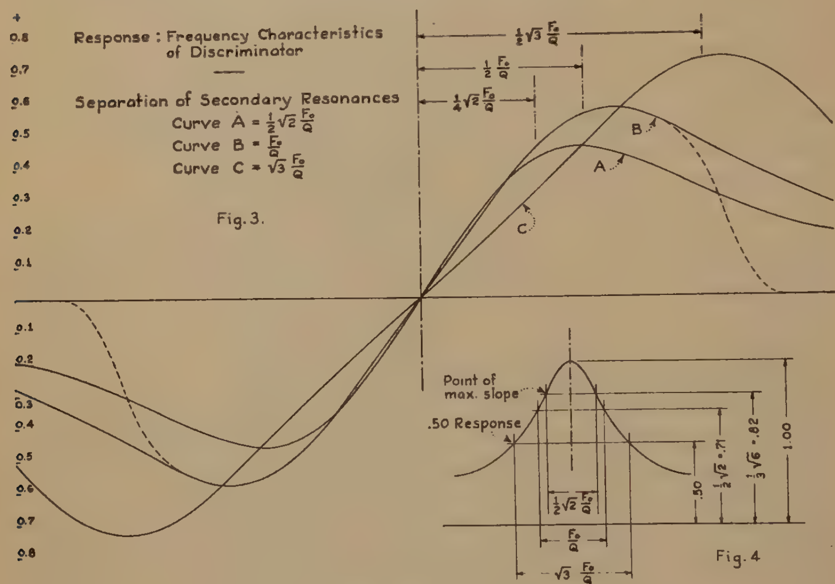
The use of two primaries in parallel in the coupling transformer is in the nature of a short cut in building a receiver in the laboratory. It was found convenient to use two ready-made intermediate-frequency transformers for the purpose, each in its own shield can and each with a pair of tuning condensers. This might still be convenient on a production basis, but it would probably be more economical in every way to use a single three-coil transformer with the primary in the center.

The circuit variation of Fig. 2 differs from that of Fig. 1 only in that a 6E7 is used as a driver, with its diodes coupled into the primary of the automatic frequency control transformer to generate automatic volume control bias. There is thus a higher carrier input into the automatic volume control rectifier than into that for automatic frequency control, which is desirable, and a considerable voltage delay may be used on the automatic volume control diodes.

One advantage of obtaining the automatic volume control bias in this manner is, that the signal modulation into the second detector is not distorted by the demodulation produced by the automatic volume control diodes, and as the second detector is relieved of its automatic volume control function the engineer is free to use any type of detector that best meets other design requirements.

The point at which the grid of the automatic frequency control driver tube is connected into the intermediate-frequency amplifier is determined by the following considerations. This driver tube must be connected to a point where the carrier strength is held nearly constant, so that in the circuit of Fig. 1 it must be connected to the input of the second detector and automatic volume control tube, as is shown schematically in the figure. If the automatic frequency control driver also drives the automatic volume control rectifier, as in Fig. 2, its putput is automatically held nearly constant no matter where it is connected, and the connection point may well be farther forward in the intermediate-frequency amplifier, at the primary or secondary of the transformer preceding the last intermediate-frequency tube. This has the merit of giving a broader channel for the automatic frequency control and the automatic volume control than for the audio detector. In either circuit, care must be taken to avoid premature overload of either the automatic frequency control rectifier or of the second detector as the case may be. In Fig. 1 the input to the grid of the automatic frequency control driver will have to be considerably attenuated, as by a capacity voltage divider as shown. In the connection of Fig. 2, the gain of the last intermediate-frequency stage will most likely have to be considerably decreased to avoid second detector overload.

When the bias output of the automatic frequency control rectifier is plotted against intermediate-frequency carrier, there results an S-shaped curve similar to those shown in Fig. 3. The theoretical form of the curve may be found by constructing two similar resonance curves, staggered by an appropriate amount, and taking the algebraic difference between their ordinates as the ordinate to the required discriminator characteristic. Actual curves are found to approximate the theoretical very closely, differing only at the two extremities where the effect of intermediate-frequency selectivity enters.



Figs. 3 and 4.

Other things being equal, maximum automatic frequency control sensitivity is obtained when this curve has the greatest slope. This occurs when the two secondaries are so tuned that points of maximum slope on their individual resonance curves are made to coincide. The point of maximum slope on a resonance curve occurs at a frequency interval from resonance equal to $\sqrt{2}\, F_0/4\, Q$, where F_0 is the resonant frequency and Q has its usual connotation of the ratio of $\omega L : R$. Hence maximum slope of the discriminator characteristic occurs when the two secondary resonant points are separated by $\sqrt{2}\, F_0/2\, Q$. However, a greater range of control is obtained when somewhat more separation is used, without greatly reducing the slope of the curve, and it seems that a separation equal to F_0/Q is about optimum.

The three curves in Fig. 3 correspond to three different amounts of separation between the secondary resonant frequencies, the particular ones chosen being shown in their relation to a single resonance curve in Fig. 4. The ordinates of Fig. 3 are shown to scale with the resonant response (rectified output) of a single secondary and diode as unity. In solid line the curves show the response of the discriminator alone, unmodified by preceding intermediate-frequency selectivity, and the dotted lines in curve *B* indicate the manner in which the preceding selectivity brings the response rapidly to zero at the edges of the band.

At 450 kilocycles an effective *Q* of 100 may be attained, representing a very good intermediate-frequency coil with a *Q* of 200 alone, reduced to half the latter value by the presence of the diode loading. Then the

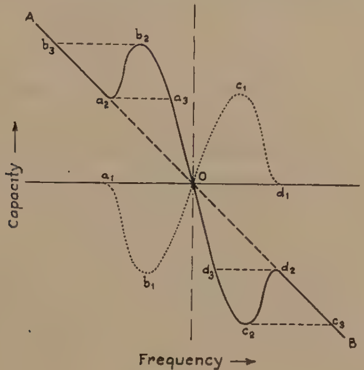


Fig. 5

optimum separation of the secondary resonance is 4.5 kilocycles, and the peaks on the corresponding characteristic (curve *B*) are about 5.4 kilocycles apart, fitting the curve nicely into the usual intermediate-frequency selectivity band. To attain the above sharpness the diode leak resistances must be made high (up to 0.5 megohm) and the coils must have less inductance than those of the usual high gain intermediate-frequency transformer (or equivalently they may be tapped).

Before considering in detail the control unit forming the remainder of the complete system, it may be well to digress briefly to explain the manner in which the system operates as a whole, in order to anticipate some of the requirements imposed upon the control circuit.

The control circuit varies the oscillator frequency, and this variation may be thought of as if due to a variation in capacity. Actually the control circuit might happen to reflect an inductance rather than a capacity across the oscillator tank, but the suggested viewpoint puts all variations of the tank circuit reactance into quantities of the same

kind. In Fig. 5 ordinates are thus taken as tank circuit capacity and abscissas as frequency, either of the oscillator or of the intermediate-frequency carrier, for the two differ by a constant amount, and frequency variations alone need be considered.

The total tank circuit capacity may be considered to consist of two parts: that part physically present as in the variable tuning condenser, and that part reflected by the control circuit. The latter is a function of the intermediate-frequency carrier and is represented by the curve $a_1\, b_1\, c_1\, d_1$ of Fig. 5. This curve is similar in shape to the characteristic of the discriminator, as shown in Fig. 3.

The relation between oscillator frequency and total tank capacity is nearly linear if only a narrow band is in question, and is expressed by the straight line AOB in Fig. 5.

A third curve, plotted with ordinates equal to those of AOB minus (algebraically) those of $a_1\, b_1\, c_1\, d_1$ expresses the capacity physically present (for physical capacity equals total minus reflected). This is the curve $A\ a_2\ b_2\ c_2\ d_2\ B$. This curve may be thought of as relating dial setting and oscillator frequency.

Where the last-named curve slopes upward from left to right, between a_2 and b_2, and between c_2 and d_2, stable operation cannot take place, for an increase in oscillator and intermediate frequency produces a decrease in total capacity through the control action, which produces a still further increase in frequency and so on. Conversely, operation is stable where the curve slopes from right to left, in the segment $b_2\, a_3\, d_3\, c_2$.

Accordingly, when a carrier is tuned in, say by decreasing the variable condenser capacity, the change in oscillator frequency is normal until the point a_2 is reached. At this point a sudden jump in frequency occurs to a_3. Thereafter operation is stable until c_2 is reached, at which point there is a second jump to c_3. In tuning in the opposite sense, starting at B and throwing in capacity, frequency jumps occur at d_2, to d_3, and at b_2, to b_3.

The ratio of the slope of $a_3\, d_3$ to that of $a_2\, d_2$ represents what may be called the sensitivity of control; it is the ratio of the amount of frequency shift due to extraneous causes that might take place without control, to that actually taking place when the control is operative. It is possible to attain a sensitivity of several hundred, so that a maladjustment that would otherwise produce an alignment error of several kilocycles will result in an actual error of only a few cycles.

For the control unit, several circuit forms have been devised, which are shown schematically in Figs. 6, 8, 9, and 10. In the present stage of automatic frequency control development in this laboratory the last

one is preferred, but the ones it has superseded are described to illustrate some of the limitations possessed by certain forms of circuit. Moreover, these circuits may have practical application within their limitations, as described more fully in the case of the one shown in Fig. 6. The chief difficulty has been to find a circuit that will function upon an oscillator tunable over a wide frequency range by means of a variable condenser.

Certain desiderata of the control circuit are the following: It is desirable that the operation of the circuit should not affect the amplitude of oscillations but should act to vary frequency only. To make for sensitivity, a large change in frequency should be produced by a small change in control bias, and this relation should hold over a sufficient frequency interval to take care of the greatest amount of unwanted mistuning or oscillator drift that will ordinarily be met with. If there

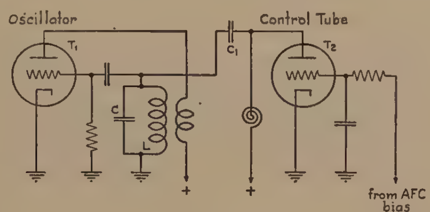


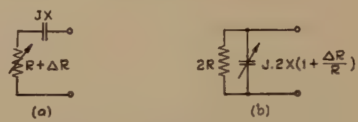
Fig. 6

is a variation in sensitivity and range over a particular wave band, it is desirable that both should be greater at the high-frequency end of the band, where mistuning and oscillator drift are more serious.

A circuit that has been very useful in experiments with frequency control is shown schematically in one of its forms in Fig. 6. For purposes of automatic frequency control as here considered this circuit has been superseded by others to be described below, but reference is made to it here because of its possible application for other related purposes where its inherent limitations are not important. For example, it forms a satisfactory frequency "wobbulator" for observing selectivity curves with the aid of the cathode-ray oscilloscope; for it can be designed to give practically linear frequency modulation, free from amplitude modulation, over a range sufficient for this purpose.

In this circuit the plate resistance of a tube is connected in series with a reactive element to the high potential side of the tank circuit. The reactive element is preferably a condenser, and is so shown in Fig. 6 as condenser C_1, but it may alternatively be an inductance, which further might take the form of the leakage inductance of a coil loosely coupled to the tank coil.

When the R_p of the tube is made numerically equal to the reactance, and varied by small increments about this carrier value, the effective shunt reactance across the tank circuit varies but the resistance remains constant. In other words, the series combination of a fixed reactance and an (approximately) equal variable resistance is equivalent to a shunt combination of a variable reactance and a fixed resistance. The equivalent forms are shown in Fig. 7.



(a) and (b) are equivalent when R = X
Fig. 7

The tank circuit is thus shunted by $2R_p$ (fixed) and an equivalent condenser $(1-\Delta R_p/R_p)C_1/2$, in which ΔR_p is the variation of the tube resistance from its center value R_p. A ten per cent change in R_p will thus change the tank circuit capacity by five per cent of C_1, which will produce a greater or smaller frequency change depending upon the value of the physical fixed tank condenser C. The variation in R_p is of course brought about by the automatic frequency control bias applied to the control grid of the tube.

This circuit has the limitation that it cannot be applied successfully to an oscillator that is to be tuned over a band of frequencies. To make use of it in a radio receiver the latter would need to have two cascaded intermediate-frequency systems to provide one nominally fixed frequency oscillator for the circuit to operate upon. Such a double intermediate-frequency superheterodyne receiver is of course a possibility, but does not appear attractive unless this design has advantages additional to permitting frequency control.

There is an additional limitation to the use of the circuit of Fig. 6. At low frequencies inconsiderable frequency variation is obtainable, and at high frequencies the output capacity of the tube acts to dilute the variations in R_p. With the usual receiving tubes the circuit is found to be limited in usefulness to frequencies ranging from about one to two megacycles.

Circuits that are better adapted to automatic frequency control make use of a radically different principle. The plate of a control tube is coupled into the high potential side of the tank circuit, and the grid of the tube is excited by a voltage ninety degrees out of phase with that appearing across the tank circuit. Several ways of obtaining this out-of-phase grid voltage are shown in the remaining figures. The plate

current in the control tube is then likewise out of phase with the tank voltage, and accordingly the tube looks like a reactance to the tank circuit. The magnitude of this effective reactance varies according to the automatic frequency control bias impressed upon the control grid of the tube, as this varies the G_m and proportionately the magnitude of the plate current.

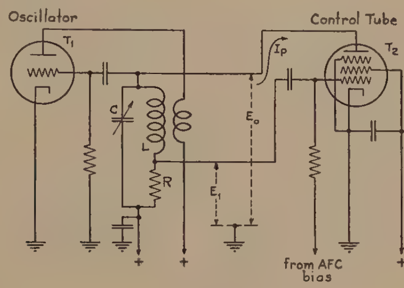


Fig. 8

One method of obtaining the out-of-phase excitation for the grid of the control tube is to take off the voltage across a resistance in series with the tank coil or condenser. This is shown schematically in Fig. 8. If E_0 is the tank voltage, the voltage across R is

$$E_1 = \frac{E_0 R}{R + j\omega L}.$$

The plate current I_p, considered as flowing into the plate of T_2 is

$$I_p = E_1 G_m = \frac{E_0 R G_m}{R + j\omega L}.$$

and the tube looks to the tank circuit like an impedance

$$Z_0 = \frac{E_0}{I_p} = \frac{R + j\omega L}{R G_m};$$

This neglects the presence of the plate resistance of T_2, which is justifiable in that T_2 will be a screen-grid tube.

The presence of Z_0 in effect multiplies the LR arm of the tank circuit by the factor $1/(1+RG_m)$ and changes the frequency in the ratio of $\sqrt{1+RG_m}$.

A practical value of R for a broadcast band oscillator is 100 ohms. With $G_m = 1000\mu a/v$, the frequency change is $\sqrt{1.1}:1$, or about five per cent. This permits a variation of ± 25 kilocycles at one megacycle which would seem ample for the present purpose.

This circuit has the merit that the tuned impedance of the tank circuit is not affected by the control action. The tuned impedance is L/RC and L and R are effectively changed by the same factor by the action of the control tube. On the other hand, the fact that L/R is appreciably constant causes the tuned impedance to vary widely when C is varied in tuning over a band, and the amplitude of oscillation varies accordingly if the oscillator has simple magnetic feedback. This possibly might be corrected by the use of series capacity feedback, but the possibilities of the circuit have not been sufficiently investigated to answer this question.

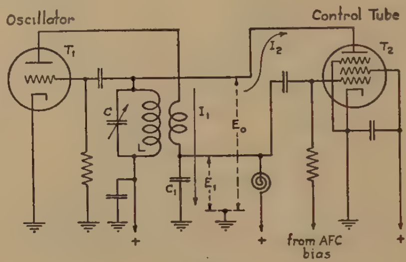


Fig. 9

It may be remarked that the series resistance R may be placed in the capacitive leg of the tank circuit instead of in the inductive leg as described. In this case the control tube will effectively appear as a resistance capacity arm across the tank circuit, and increase in the G_m of the control tube will lower the oscillator frequency. The action of the circuit is otherwise quite the same as that in which the resistance is in the inductive leg.

Another method of obtaining the out-of-phase grid voltage is the following: If the plate impedance of the oscillator tube is very high, and stray admittances from plate to cathode are negligible, the current flowing in the external plate circuit is in phase with the tank circuit voltage. A reactive element in the plate circuit will therefore develop an out-of-phase voltage which may be impressed upon the grid of the control tube. A circuit utilizing this principle is shown in Fig. 9, in which the reactive element is the condenser C_1 in series with the oscillator tickler winding.

To develop the theory of this circuit, let the mutual conductance of the oscillator tube T_1 be G_1 and that of the control tube T_2 be G_2; and let the reactive coupling element be generalized as an impedance Z_1. If the tank circuit voltage is E_0, the current I_1 flowing *out* of the plate T_1 is

$$I_1 = -E_0G_1$$

and the voltage E_1 across Z_1 is

$$E_1 = -E_0 G_1 Z_1.$$

The plate current I_2 in the second tube T_2, flowing out of the tank circuit and into the plate, is

$$I_2 = +E_1 G_2 \doteq -E_0 G_1 G_2 Z_1$$

and the tube appears as an impedance

$$Z_0 \doteq \frac{E_0}{I_2} = -\frac{1}{Z_1}\left(\frac{1}{G_1 G_2}\right).$$

The effective impedance of the plate of T_2 is then the negative reciprocal of the coupling impedance, with respect to a resistance product $1/G_1G_2$. A condenser such as C_1 thus reflects as a negative inductance $-C_1/G_1G_2$; a positive inductance reflects as a negative capacity; and a negative inductance, which may be realized by using the mutual of a transformer for Z_1, reflects as a positive capacity. Any resistive element in Z_1 reflects as a negative resistance or conductance and tends to aid the feedback of the oscillator tube T_1.

It seems preferable to make use of a condenser for Z_1 for this causes the control tube to give inductance variation and realizes a range and sensitivity of control which are proportional to frequency over each particular tuning band. If an inductance were used for Z_1 the circuit would give a capacity variation and range and sensitivity would be proportional to the cube of frequency.

In the idealized case this circuit would be capable of changing frequency from an original maximum determined by the physical L and C of the tank circuit, down to zero frequency, for the reflected negative inductance in shunt to that physically present may indefinitely increase the total inductance. In practice an extremely wide frequency coverage is obtainable, thirty per cent or more of the center frequency.

It has been found however, that the circuit does not fulfill expectations over the whole wave band covered by the tuning dial. The assumption that the whole oscillator plate current flows through C_1, i.e., that the plate impedance is high and that there are no strays to ground, is of course only very roughly tenable. Actually the coupling impedance is not merely C_1 but instead is a complicated network, involving the tank circuit reflected by the tickler coil, the leakage inductance of the latter, etc. Unforeseen resonances probably occur; what is observed is that at the high-frequency end of the band the system breaks into violent self-maintained frequency modulation. Apparently the two tubes alternate in the role of oscillator, and the generated frequency is different for the two modes.

This effect can be avoided by making C_1 larger, or reducing the maximum value of G_2, either acting to reduce sensitivity. A more logical method is to decouple condenser C_1 from the tank circuit by using an electronic coupled oscillator; this gives rise to the circuit shown in Fig. 10.

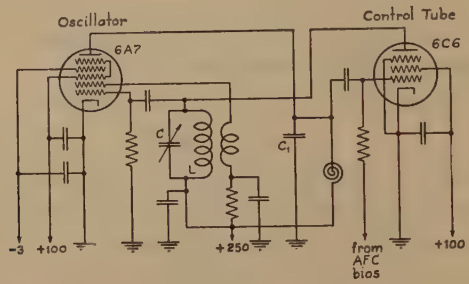


Fig. 10

Here the oscillator is a 6A7 tube, with normal electrode voltages. Grid 4 is not used, and is by-passed to cathode and given a fixed bias of -3 volts. Condenser C_1 coupling to the grid of T_2 is placed in the plate circuit of the 6A7, with parallel feed around it through a radio-frequency choke. The mutual G_1 of the above analysis is now $G_m\ (1-p1)$, and, if desired, frequency variation could be brought about by changing the bias on grid 4, as well as that on the control grid of T_2.

It is conceivable that with proper choice of constants, C_1 and the parallel choke might be replaced by the primary of an intermediate-frequency transformer, and a radio-frequency signal introduced upon grid 4, thus permitting the 6A7 to act as a combined converter and controlled oscillator. It is apparent however, that automatic volume control bias could not be used upon the signal grid, which is a disadvantage in the usual case where large antenna inputs must occasionally be handled. For this reason it seems preferable to use a separate oscillator and converter.

A variant of Fig. 10, (not illustrated), consists in the substitution of an ordinary screen-grid tube for the 6A7, using the screen for the oscillator anode and the plate for the electronically coupled output anode.

In Figs. 8 to 10 the oscillator and control circuits only have been diagramed, omitting the coupling to the mixer tube. This latter may take the form of an extra winding in the cathode circuit of the mixer, or perhaps a better way is to use a 6A7 for the mixer, with No. 1 grid connected directly to the grid of the oscillator. In this case No. 2 grid may be connected to cathode. With this connection it will probably be found that band coverage suffers from the high minimum capacity of the

oscillator circuit, for the tank condenser will be paralleled by the input capacity of the mixer tube and the output capacity of the control tube in addition to the usual oscillator input capacity. To decrease the effect of this additional stray capacity, the mixer grid may be connected to the control tube plate through a blocking condenser, with a biasing leak, and the common junction connected to a close coupled secondary on the tank circuit, which will also serve to supply direct current to the control tube plate. This circuit arrangement reduces control sensitivity in proportion to the impedance ratio of the two windings.

In any of the control circuits of Figs. 8, 9, and 10, the control tube can be a sharp cutoff radio-frequency pentode such as the 6C6. When the automatic frequency control bias increases in the positive direction, a limiting point is reached in which the control tube biases itself by grid current through the filter resistor system, the bias attained being nearly equal to the peak value of the impressed oscillator frequency voltage. As the tube might be called upon to run for long periods in nearly this condition, the screen voltage must be reduced to avoid excessive plate current and consequent shortness of life. For bias variation in the negative direction, control action is limited by plate current cutoff. With sufficient input to the automatic frequency control rectifier, the total range of control is thus independent of the maximum value of the bias generated. If the input is in excess of this amount, the sensitivity is increased, but at the same time it will be necessary to bring the signal nearer to exact alignment before automatic frequency control will "take hold."

To take up more general considerations, it can be said that an automatic frequency control system can be designed and built, using the described circuits or variations thereof, that will operate in the anticipated manner. It is believed that an important step has been taken in the development of a commercially practicable system, but undoubtedly a number of problems remain to be solved. Automatic frequency control has not yet reached the point where it can be installed in an existing model of receiver merely because there happens to be room for three more tubes and a few necessary parts; the whole receiver will have to be redesigned from the ground up. It is further very evident that if and when the system is placed upon a commercial basis, it will apply at first only to receivers in the higher price class, where the selling price is determined by the cost and not vice versa.

It is inherent in a control system like the present that the signal must be amplified and selected before it can operate a control. Hence the automatic frequency control cannot be expected to bring in a signal that lies outside the selectivity band; the signal must be brought in by

manual or mechanical means, but thereafter the automatic frequency control can be expected to maintain it in alignment. A possible difficulty is to be anticipated in this connection. If control is momentarily lost and the physical tuning at the time is half a channel or more in error, the signal will disappear and will not return when control is regained. This might well happen in short-wave reception if severe fading of the carrier occurs. It is possible that a signal fading so badly would be of little use to the listener in any event, and further if the control could be voluntarily rendered inoperative—there would certainly be some means provided for this—the receiver would be no worse than one not equipped with control. This phase of the question does not appear to be serious.

It will be necessary to provide means for removing a station signal from the band after it has been once tuned in, to make way for another. If signals are closely spaced the mistuning sufficient to take an incumbent signal out of control will be sufficient to carry a number of adjacent weaker ones across the band in its wake. To avoid this the control must be rendered inoperative during the process of selecting a station. This had best be done automatically, probably by a special switch incorporated in the tuning knob, so that the contact of the operator's hand upon the latter will short-out the automatic frequency control bias. Tuning would then be performed with sufficient accuracy to bring the signal into the band, and automatic frequency control would come into play and complete the tuning when the hand is removed from the knob.

It is to be anticipated that stability of alignment will be an important practical consideration. In regard to the alignment of the discriminator circuits with the intermediate-frequency channel no serious difficulties are expected, for in the class of receiver to which automatic frequency control seems adapted there will most probably be two intermediate-frequency stages (three transformers), high gain per stage will be unnecessary, and high capacity circuits will be used. Changing tubes or readjusting the position of grid leads will therefore not affect the tuning of the circuits to any great extent. Tuning drift due to thermal causes will be small (when measured in cycles) even at an intermediate frequency as high as 465 kilocycles, and if the circuits drift in the same direction the net effect will be further reduced.

Another problem that will require thorough investigation is that of the effect of change of the control tube characteristics with variation in line voltage and during the life of the tube. An assumed concrete example will illustrate the possible nature of this problem.

Assume that the sensitivity is such that control action is capable of producing a ten per cent change in the oscillator frequency. The initial

automatic frequency control bias will then be chosen so that this ten per cent is divided equally above and below the operating point, and the receiver lined up accordingly. Then if the control tube is cut off completely a five per cent shift will occur, and any decrease in the G_m of the tube will produce a proportionate shift. If the G_m decreases say twenty per cent during life, the shift of the operating point will be twenty per cent of five per cent or one per cent of the base frequency, equal to ten kilocycles at one megacycle. This change will not affect the automatic centering of the signal in the intermediate-frequency band, but it will affect the tracking of the oscillator with the radio-frequency circuits, and with the frequency indications upon the dial.

The error can be very easily corrected by adjusting the initial automatic frequency control bias, but this would not be desirable in commercial practice and is to be avoided. It is thought however, that the difficulty can be minimized to the vanishing point by operating the tube with conservative voltages, to insure that its life will exceed that of the other tubes of the receiver, and in addition by making use of self-bias upon the tube, which will very much diminish the change in characteristics. This self-bias will also reduce the response of the control tube to the automatic frequency control bias, but the sensitivity can be made up by increasing the intermediate-frequency input to the automatic frequency control rectifier.

Experimental work on automatic frequency control in this laboratory is thus proceeding on the assumption that the difficulty mentioned will be negligible when the above precautions are taken. If it turns out that this is not the case, a "main strength" remedy is at hand in the use of a differential control circuit. Such a circuit would be merely the push-pull variant of the single side circuit described above, and the additional complexity would be by no means prohibitive if no better way of meeting the problem is found satisfactory.

Attention must be paid to the relative time delays of the automatic frequency control bias filter system and that of the automatic volume control system. In general, the automatic frequency control system must be the slower of the two. Otherwise a signal will be brought into the center of the band, where gain is high, more rapidly than the gain can be reduced by automatic volume control action, and momentary blasting will result. On the other hand, the automatic frequency control system should have a smaller time constant than the power supply unit, in order that it may follow and prevent changes in oscillator frequency due to changes in line voltage.

Proceedings of the Institute of Radio Engineers
Volume 23, Number 10 *October, 1935*

RECENT DEVELOPMENTS IN MINIATURE TUBES*

BY

BERNARD SALZBERG AND D. G. BURNSIDE
(RCA Manufacturing Company, RCA Radiotron Division, Harrison, New Jersey)

Summary—*The development of two indirectly heated miniature tubes, a triode and a sharp cut-off amplifier pentode especially suited for use at high frequencies, is described. The electrical and mechanical factors involved in the design and application of these tubes are discussed and their novel structural appearance is described.*

Because of their decreased lead impedances, interelectrode capacitances, and transit times, these miniature tubes allow considerable improvement to be made in high-frequency receiving equipment. It is possible to operate the triode as an oscillator in a conventional circuit down to a wavelength of approximately 40 centimeters. The pentode can be operated as a radio-frequency amplifier down to a wavelength of approximately 70 centimeteres. It is practicable to obtain stable gains with it of from ten to fifteen at three meters, a wavelength at which standard tubes are almost entirely ineffectual. Both tubes can be used, down to much lower wavelengths, in exactly the same manner and for the same applications that the corresponding conventional tubes are used; i.e., as oscillators, amplifiers, detectors, converters, and as negative-resistance devices.

The small size of the tubes and their novel structural design allow compact and convenient receiving equipment to be built. Even at the longer wavelengths, they are applicable to a large number of uses for which their excellent characteristics, small size, and low weight make them particularly useful.

INTRODUCTION

EARLY work on the extension of the high-frequency limit of receiving equipment, which made use of conventional radio-frequency amplifier circuits built up around standard tubes, indicated that the amplifier section of the receiver became less and less effective as the signal frequency was increased and that ultimately, at frequencies of the order of one hundred megacycles, the amplifier was virtually useless. Similarly, the detector section and the oscillator section (when one was employed) of the receiver became more and more ineffectual as the frequency was raised and although they continued to operate beyond the limiting frequency of the amplifier, they too ultimately became inoperative. Improvements in the circuits proper resulted in only relatively slight improvements in the over-all performance of the receiver, so that it became apparent that the tubes themselves limited the operation of the equipment.

Recourse was then possible only to positive-grid operation of the available tubes, either in the oscillating state (Barkhausen-Kurz or

* Decimal classification: R330. Original manuscript received by the Institute, April 26, 1935. Presented before New York meeting, March 6, 1935.

Gill-Morrell), or in the nonoscillating state (Hollmann or Carrara). These methods were not entirely satisfactory, however. With them, no cascaded amplification was obtainable at the carrier frequency, operation was invariably attended by high internal tube noise, over-all sensitivity was poor, careful adjustments were required, and the equipment was generally unreliable, at least for anything more than experimental use.

B. J. Thompson realized that the limitations of the conventional tubes were the result of their size. G. M. Rose, Jr. and he built small tubes and with them conclusively demonstrated that the limitations to the successful operation of vacuum tubes at the higher frequencies may be overcome by reducing the dimensions of the tubes.[1,2]

The possibilities of such tubes aroused an interest sufficiently widespread to warrant further development, and the "acorn" tubes are the present results of this development work.[3]

Factors Involved in the Design of Miniature Tubes

The essential principle upon which these tubes are based is the Model Theorem, or the Principle of Similitude. It is possible to show from the fundamental differential equations and the boundary conditions involved that if all of the linear dimensions of a tube structure are divided by a constant factor, say n, then the electrode currents, transconductance, amplification factors, and plate resistance will remain substantially constant, but the lead inductances and capacitances, the tube capacitances, and the time of passage of the electrons between the various electrodes will be divided by n. The direct-current lead resistances will be multiplied by n, but the alternating-current lead resistances will ordinarily be increased by something less than this factor. The allowable plate dissipation and the available emission of the tube, however, will be divided by n^2, and the current densities will be multiplied by n^2. The latter considerations are important in any application of this principle to power amplifier or transmitter tubes. Physically, the tube will be reduced in its over-all dimensions by a factor n, and its weight by a factor n^3.

For mechanical reasons, however, it is not feasible to reduce all of of the tube dimensions to the same degree. For example, a reduction by a factor of four of all of the linear dimensions of the type 56 tube would require grid side rods of $6\frac{1}{4}$ mils diameter, grid wire of 0.83 mil

[1] B. J. Thompson, "Tubes to fit the wavelength," *Electronics*, August, (1933).
[2] B. J. Thompson and G. M. Rose, Jr. "Vacuum tubes of small dimensions for use at extremely high frequencies," Proc. I.R.E., vol. 21, pp. 1707–1721; December, (1933).
[3] Bernard Salzberg, "Design and use of 'acorn' tubes for ultra-high frequencies," *Electronics*, September, (1934).

diameter, a cathode sleeve of $12\frac{1}{2}$ mils diameter, and a cathode coating of approximately 0.75 mil thickness. Parts of such minute dimensions would be extremely difficult to manufacture with present-day equipment. Consequently, as is usual in many engineering problems, it is necessary to arrive at a practical solution which will give the essential results.

In addition to the limitations imposed upon design by consideration of the internal structure, a number of electrical requirements must be considered in the design of such miniature tubes. For example, the tubes should not only be suitable for high-frequency work, but because their reduced size and weight are frequently advantageous at lower frequencies, it is desirable that they be suitable for practically any use to which the standard tubes can be put. They should be adaptable to either battery or alternating-current operation, and should require a minimum of heater power. Their electrode voltage and current ratings should be such as to fit, in so far as possible, existing auxiliary circuit equipment. A reasonable life expectation and good stability of characteristics during life are also essential requirements. As regards external structure, it is imperative that the basing designs be such as to permit the aging and testing operations incidental to the manufacture of the tubes to be carried out conveniently, and to allow easy insertion of the tube in circuit equipment without introducing undue losses. The lead terminals should be placed so that they permit short connections between the tubes and the associated circuits. Finally, it is desirable that the tube designs be suitable for economical production maintainable within reasonably close electrical limits.

Considerations such as these circumscribe the design of the tubes. Their static characteristics must be set by a study of the various uses to which the tubes might be put and by a careful weighing of their relative importance. For the triode, such considerations indicated the desirability of obtaining a plate resistance of the order of ten to fifteen kilohms and an amplification factor above fifteen, together with a ratio of grid-plate transconductance to plate current as high as possible. For the sharp cut-off pentode, it was desirable to obtain under amplifier operating conditions a plate resistance of over one megohm and a transconductance of at least 1000 micromhos.

To effect a significant improvement in high-frequency behavior, and yet have tubes which could be manufactured, it was decided to reduce the electrode spacings to minimum distances of the order of five mils. This reduction would extend the upper frequency limit of operation beyond that of existing commercial tubes by a factor of approximately four, and at the same time would keep the design of the tubes within the realm of attainable manufacturing technique.

To make the tubes suitable for general battery and alternating-current use, the heater voltage was set at 6.3 volts. It was decided to use a conventional cylindrical cathode of the smallest practical diameter in order to realize the minimum heater power. Its length was so chosen that the desired static characteristics could be obtained with low interelectrode capacitances. Maximum plate voltages of 180 volts for the triode and 250 volts for the pentode were chosen as consistent with the requirements of low transit time and good life performance.

To obtain short leads, low losses, suitable means for effective by-passing at the higher frequencies, and a practical form of base, it was necessary to abandon the conventional form of pinch stem and standard base assembly. Instead, the tube elements were connected to heavy leads suitably positioned and fastened to mica spacers; these leads were then sealed in and used as base pins.

Except for transit-time considerations, the actual calculations involved in the design of these tubes are similar in all respects to the usual tube structure calculations. These are of a semiempirical nature, based in part upon known derivations of the space-charge-limited current between the elements of diodes of simple geometrical shape, and in part upon approximate analyses of the electrical field of tubes of ideal configurations.[4,5,6,7] As regards the transit time between the various electrodes, it is important for our purposes to keep the distances traversed by the electrons as short as possible, and to get them moving as fast as possible (by making the effective potentials in the planes of the various electrodes as high as possible).

Description of the Tubes

The tubes are shown in Fig. 1, and their size can be judged from the size of the golf ball which is included in this group photograph. The remarkable resemblance of the triode to an acorn, both in size and shape, has resulted in the trade designation for these tubes as the "acorn" series. One of the practical difficulties which arose as a result of the small size of the tubes was the difficulty of etching the usual trade name and type number on the bulb.

[4] I. Langmuir and K. T. Compton, "Electrical discharges in gases. Part II, Fundamental phenomena in electrical discharges," *Rev. Mod. Phys.*, vol. 3, p. 191; April, (1931).
[5] F. B. Vogdes and F. R. Elder, "Formulas for the amplification constant for three-element tubes in which the diameter of grid wires is large compared to the spacing," *Phys. Rev.*, vol. 24, p. 683, (1924).
[6] Bernard Salzberg, "Application of conformal transformation to some problems in vacuum tube engineering," Thesis for M.E.E. degree, Polytechnic Institute of Brooklyn, 1933.
[7] S. Koizumi, "On the amplification constants of multi-electrode tubes," *Jour. I. E. E.* (Japan), Abstracts, vol. 10, p. 18, (1934).

The internal structures of the tubes are shown in Figs. 2 and 3. Both tubes are provided with a cylindrical cathode, inside of which is inserted a folded type of insulated heater. Fig. 4 illustrates the relative dimensions of this heater-cathode assembly and a common variety of pin. This gives some idea of the size of the parts and the precision and care required in the fabrication of such tubes. The No. 1 grids are elliptical in order to reduce the otherwise severe control effects of the side rods. The suppressor grid of the pentode is particularly useful in reducing the reaction between control grid and plate. The various elements are spaced very closely along the minor axis of the grids—where the cathode emission is most useful—in order to reduce the

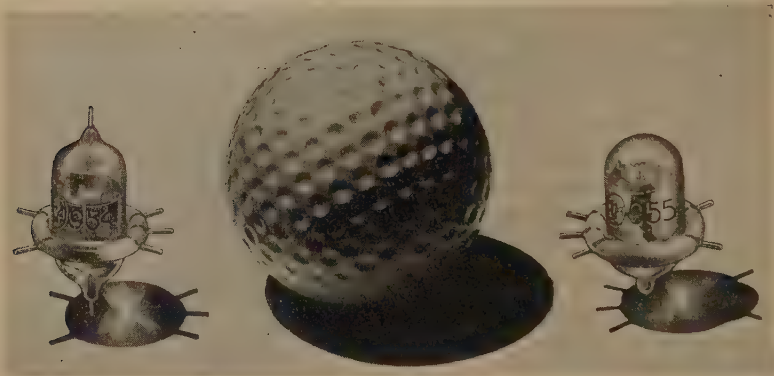


Fig. 1—Miniature triode and pentode compared with golf ball.

transit time and increase the transconductance, and somewhat less closely along the major axis in order to decrease the capacitances and increase the structural strength.

In both tubes, two accurately punched mica spacers serve to hold the elements in position. In the triode, the whole assembly is fastened in place by means of small lugs which are bent out from the ends of the plate. In the pentode, an additional mica spacer is used to insulate the plate lugs from the shielding structure, the latter being held in place by means of two support rods. In both tubes, one of the mica spacers is used as a sort of a stem, the heavy base pins and the lighter leads to the electrodes being fastened around its periphery. The "getter" material is enclosed within a flat tab welded, in the triode, to one of the plate lugs, and in the pentode, to the shielding structure.

The mount, in each case, is placed within a bulb which consists initially of two cup-shaped heavy preformed glass sections, the shallower one of which has an exhaust pipe attached to it. The main seal,

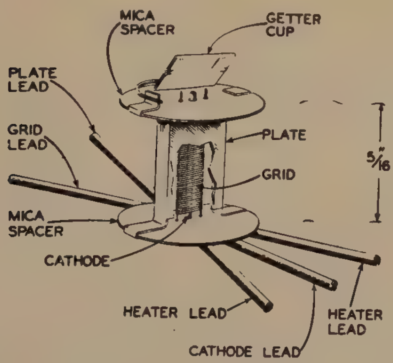


Fig. 2—Internal Structure of the triode.

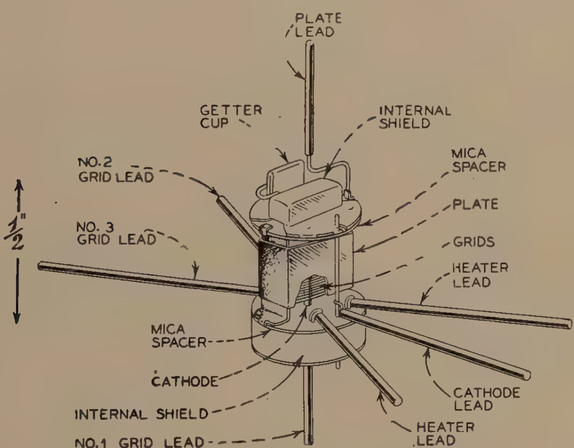


Fig. 3—Internal structure of the pentode.

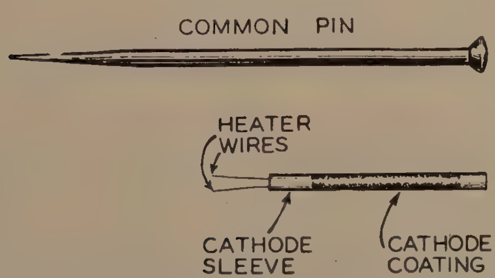


Fig. 4—Comparison of heater cathode assembly and an ordinary pin.

and in the case of the triode the only seal, is of the joined-flare type, made at the plane of the tube leads between the two glass sections. After the sealing-in operation, these leads are cut quite short, the com-

bination of the heavy glass of the joined-flare seal and the stub pins constituting a practical and extremely sturdy base. This arrangement obviates the need for soldering the tube to the circuit elements and also avoids the deleterious effects of the conventional base. The leads themselves are radial, the two heater leads being symmetrically placed 30 degrees on either side of the cathode lead, the grid and plate leads (in the triode), 60 degrees apart on the same diameter as the heater leads. In the pentode, the latter two leads are replaced by the numbers 3 and 2 grid leads, respectively, and the plate and control-grid leads are brought out at opposite ends of the bulb, a very convenient circuit arrangement. These lead arrangements provide the necessary separation between the active radio-frequency leads and the grounded radio-frequency leads. In the pentode, the control-grid lead is brought out through the exhaust pipe, an innovation in tube manufacture: incidentally, the resultant tube is thus a very special form of the "tipless" variety. The getter flash in this tube is of the directed beam type, necessary because of the small size of the structure in order to avoid leakage between elements. The inside surface of the pentode bulb is coated with carbon to reduce the emission of secondary electrons which results from the bombardment of this surface by primary electrons which escape through the interstices between the mica spacer and plate. This is particularly important at the lower frequencies, where it is desirable to keep the high load impedances from being shunted by the additional plate losses introduced by such effects.

Characteristics of the Tubes

I. Static and Low-Frequency Characteristics

The heaters of both tubes are rated at 6.3 volts, making them suitable for both battery and alternating-current operation. The tubes themselves were designed so that the electrode voltages could be set at "preferred" values.

A typical plate family for the triode is shown in Fig. 5. Because of the reduced allowable plate dissipation, such miniature tubes are not especially suited for use as audio power amplifiers. Within maximum ratings, however, this tube is entirely suitable for audio-frequency amplifier and high-frequency oscillator uses. Operated as a class A audio amplifier with a load of approximately 19,000 ohms, the triode delivers an undistorted power output of 130 milliwatts. Table I shows the comparative characteristics of the miniature triode and a standard triode, type 76. It will be observed that at the same electrode voltages, not only is the miniature triode not *inferior* to the conventional tube,

but it is actually *superior* to it. One point not specifically tabulated is the electron grid current which flows from the grid due to thermionic emission. This is negligible, even with abnormal heater voltages, due to a combination of methods which reduce the temperature of the grid and otherwise reduce the tendency of the grid to emit. The lower

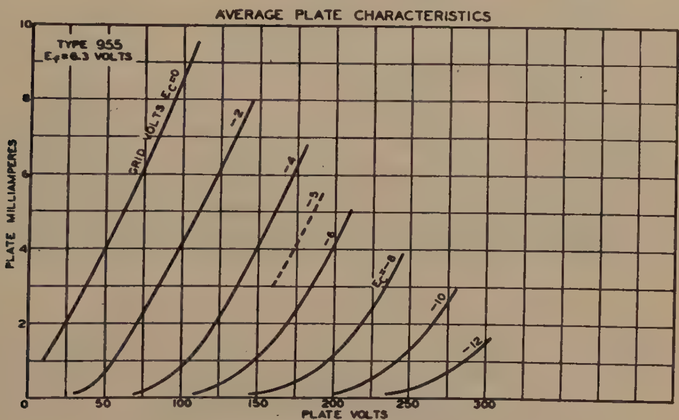


Fig. 5—A typical plate family for the triode.

capacitances of the miniature tube are very desirable at the lower frequencies, as well as the higher frequencies, because they allow higher gain over wider frequency spans. Laboratory tests of tubes of this type made up to the present writing have indicated over 1000 hours of continuous and useful life.

TABLE I
COMPARATIVE CHARACTERISTICS

	RCA-76 Conventional Triode	RCA-955 Acorn Triode	
Heater Voltage	6.3	6.3	volts
Heater Current	0.30	0.15	amp
Plate Voltage	180	180	volts
Grid Voltage	−9.8	−5.0	volts
Plate Current	3.2	4.5	ma
Amplification Factor	13.8	25	
Plate Resistance	11200	12500	ohms
Grid-Plate Transconductance	1200	2000	μmhos
Input Capacitance	3.5	1.0	μμf
Output Capacitance	2.5	0.6	μμf
Grid-Plate Capacitance	2.8	1.4	μμf

Figs. 6 and 7 illustrate typical plate and transfer families, respectively, for the miniature pentode. It will be observed that these static characteristics are similar to the usual characteristics obtained with conventional tubes of this type; i.e., high grid-plate transconductance,

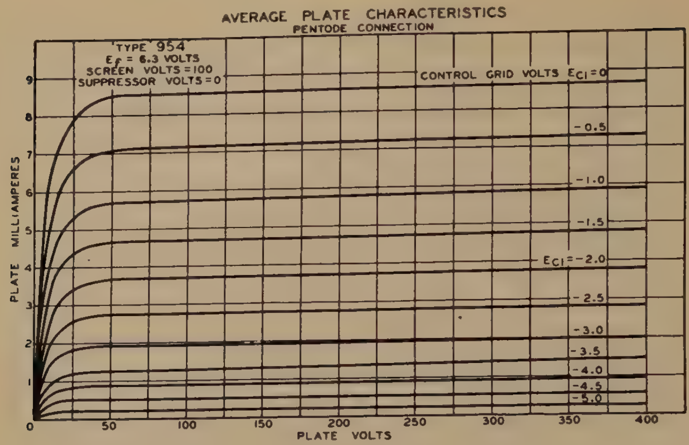


Fig. 6—A typical plate family for the pentode.

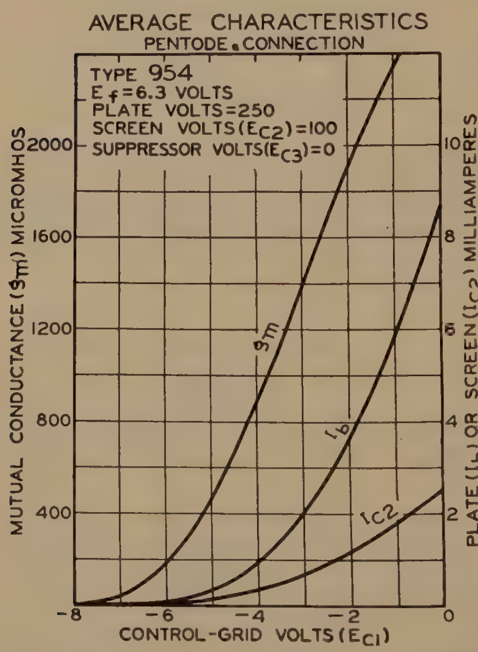


Fig. 7—A typical control-grid plate transfer family for the pentode.

high plate resistance, and sharp cutoff. Table II shows the comparative characteristics of the miniature pentode and a standard pentode, type 6C6. It will be observed, once more, that at the same electrode

voltages, the miniature pentode is actually superior to the conventional pentode, even on a strictly static comparison basis.

TABLE II

Comparative Characteristics

	RCA-6C6 Conventional Pentode	954 Acorn Pentode	
Heater Voltage	6.3	6.3	volts
Heater Current	0.30	0.15	amp
Plate Voltage	250	250	volts
Screen Voltage	100	100	volts
Suppressor Voltage	0	0	volts
Control-Grid Voltage	−3.0	−3.0	volts
Plate Current	2.1	2.0	ma
Plate Resistance	1.5	1.5	megohms
Control-Grid—Plate Transconductance	1225	1400	μmhos
Input Capacitance	5.0	3.0	μμf
Output Capacitance	6.5	3.0	μμf
Control-Grid—Plate Capacitance	0.010	0.005	μμf

II. High-Frequency Characteristics

The reduction in size of these tubes has allowed operation of the triode as a radio-frequency feed-back oscillator in conventional circuit arrangements down to a wavelength somewhat below 40 centimeters, and the operation of the pentode as a radio-frequency amplifier, also in conventional circuit arrangements, down to wavelengths of the order of 70 centimeters.

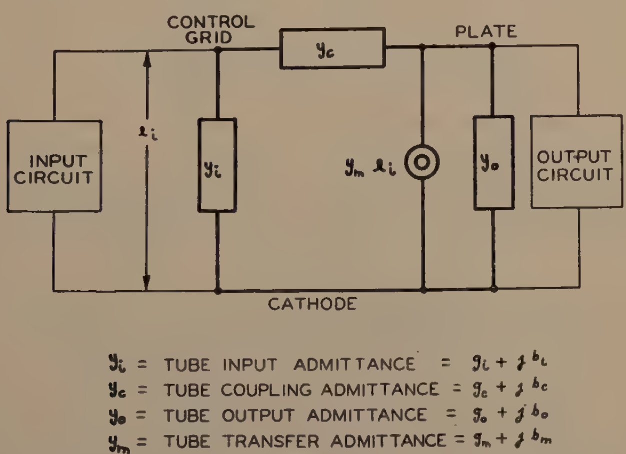


Fig. 8—The schematic first order equivalent amplifier circuit representation of an amplifier tube.

Fig. 8 shows the schematic first order equivalent amplifier circuit representation of an amplifier tube. In order to specify completely the

behavior of such a tube at any frequency, it is necessary to specify the input, output, coupling, and transfer admittances at this frequency. At low frequencies, the input and coupling admittances are essentially pure susceptances, capacitive in nature. The output admittance consists of the plate conductance, usually in a radio-frequency amplifier pentode quite small in magnitude, and a susceptance due to the output capacitance. The transfer admittance consists of only a real part, the familiar g_m.

At high frequencies, the input admittance, and (unless the tube is well screened electrically) also the coupling admittance, acquire a real part and an altered imaginary part. In addition, the output admittance acquires an increased conductance and an altered susceptance. The transfer admittance also acquires an imaginary part. The increase in the conductance components of the input, output, and coupling admittances tends to reduce the gain which can be built up by the amplifier, and ultimately to limit its effectiveness. The change in the susceptance components of these admittances alters the tuning of the associated circuits, and in the case of an oscillator provides a potential form of frequency variation. The phase shift in the transfer admittance makes conditions for stability more favorable, but also affects the gain.[8] These admittances are functions of the time of passage of the electrons between the various electrodes. B. J. Thompson and W. R. Ferris of the RCA Radiotron Laboratory have already discussed the loading effects due to the increase of the input conductance of tubes at high frequencies,[9] and W. R. Ferris and D. O. North, also of the same laboratory, have made experimental and theoretical studies of this effect.[10]

Typical results of measurements of the equivalent shunt input resistance of the "acorn" pentode and the RCA-6C6 are shown in Fig. 9. This is the component of resistance which is due to the actual electronic flow past the control grid. The heavy section of the curve indicates the range over which measurements were made. The light line continuations are extrapolations. There is some justification for this extrapolation procedure because theory indicates that for small values of transit angle the conductance varies as the square of the frequency. The curves indicate that even at thirty meters, where it is perfectly feasible to build up a resonant circuit impedance of 150,000 ohms, the shunting effect of the conventional tubes is serious. At three meters, it

[8] Bernard Salzberg, "Notes on the theory of the single stage amplifier," to be published.

[9] B. J. Thompson and W. R. Ferris, "Grid circuit losses in vacuum tubes at very high frequencies," presented before joint U.R.S.I.–I.R.E. meeting, Washington, D. C. April 27, 1934; Abstract published Proc. I.R.E., vol. 22, p. 683; June, (1934).

[10] To be published.

amounts for conventional tubes to approximately 1600 ohms; this explains why the gain is so low at this wavelength when standard tubes are used.

In addition to the shunting which is due to the finite time of passage of the electrons between the various electrodes, there is an additional loss incurred as a result of the dielectric losses in the tube-lead support —glass, socket, etc.—and as a result of the capacitive currents which flow by way of the lead wires through the tube capacitances. The latter

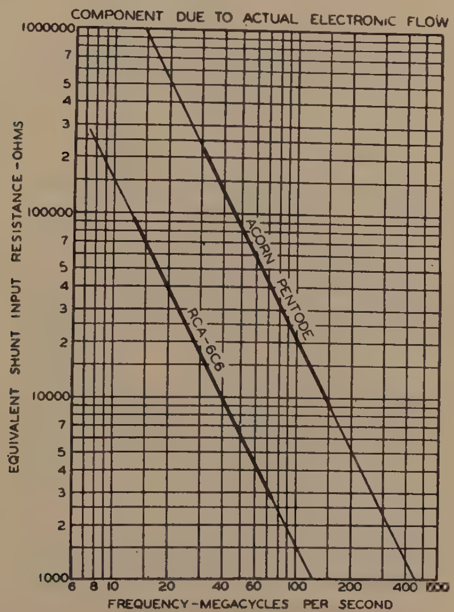


Fig. 9—The equivalent shunt input resistances of the "acorn" pentode and the RCA-6C6 as a function of the frequency. This is the component of resistance which is due to the actual electronic flow past the control grid.

cause assumes importance at the high frequencies, as shown by the curves of Fig. 10. Here is shown the shunting which takes place in the output circuit of a standard tube and of an early developmental "acorn" pentode which was used to study this effect. The losses are due partly to the dielectric loss, but mainly to the capacitive currents which passed through the leads. The same leads and by-pass condensers were used for both tubes. These curves are shown for the purpose of illustrating how significant this effect may become, and therefore how important it is to keep the radio-frequency losses in the connecting wires and the by-pass capacitors as low as possible. As seen from the

curves, this loss at the higher frequencies completely overshadows the shunting effect of the static or low-frequency plate resistance.

The measurements of the input and output admittances of tubes were made by a substitution method. A resonant circuit was coupled to a high-frequency oscillator and the voltage developed across this circuit at resonance was measured by means of an "acorn" pentode voltmeter, of the type shown in Fig. 13. The tube was operated at small space currents to reduce the electronic loading effects to a minimum.

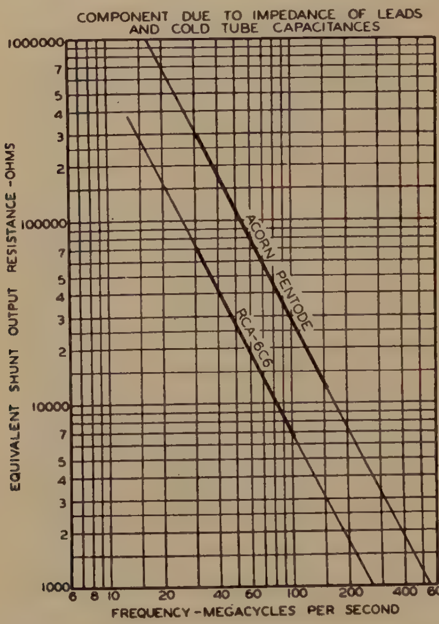


Fig. 10—The equivalent shunt output resistance of an early developmental "acorn" pentode and the RCA-6C6. This is the component of resistance which is due to the cold resistance of the tubes and the associated connecting wires.

This measurement circuit was calibrated by placing resistors, whose value at the operating frequency had been checked previously by a reactance substitution method, across the resonating capacitor, retuning to resonance and plotting the resulting resonant voltage against resistance. The conductance component of the admittance of a tube under test was then indicated by the value of the resonant voltage. The susceptance component was given by the frequency and the change (of the tuning capacitor) required to establish resonance.

The effects of the tube conductances on the associated input and output circuits may be minimized by matching the tube impedances

to these circuits. This may be done in a variety of ways, perhaps the simplest of which consists merely in connecting the tube at an intermediate point on the coil. It should be emphasized that since the shunt-

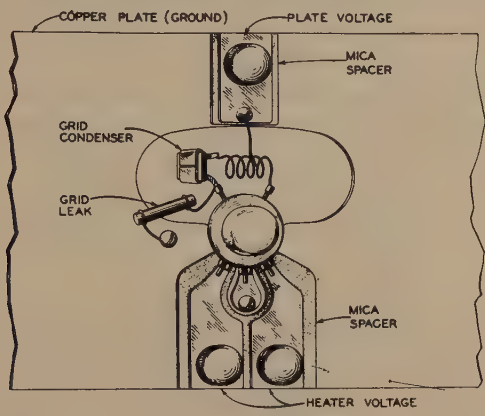


Fig. 11—A typical triode oscillator set-up.

ing effects are considerably less for the miniature tubes, there is less need for circuit arrangements of this type for these tubes. However, even for them, it becomes desirable ultimately to make use of such schemes.

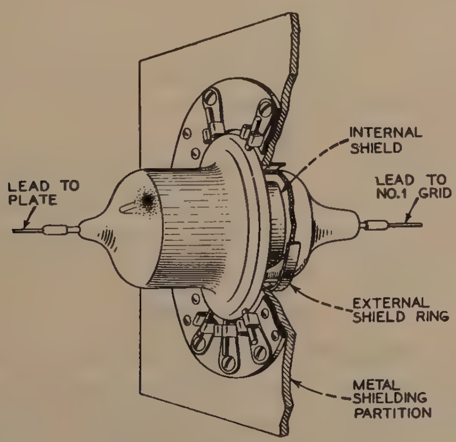


Fig. 12—A typical shielding arrangement for the miniature pentode.

Field measurements of noise developed in tubes at high frequencies indicate that the miniature tubes are definitely superior to the conventional tubes. Although elementary theory indicates that the shot noise in the plate circuit of a tube decreases with increasing interelectrodic

transit angle, the initiation of a noise component in the grid circuit which is a function of the cathode-grid and of the grid-screen transit angle results in an increase in the over-all noise.[11]

Figs. 11 and 12 illustrate how conveniently these tubes fit in with high-frequency circuit requirements. In the triode, the leads are arranged so that the active radio-frequency leads are on the opposite side of the flare seal from the radio-frequency grounded leads. In the pentode, the grounded radio-frequency leads are all brought out in the

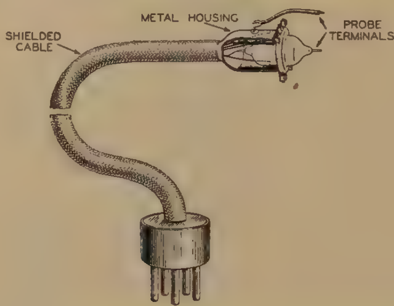


Fig. 13—A probe voltmeter using a miniature pentode.

plane of the flare seal, so that this may be used as a ground reference plane. The two active radio-frequency leads are brought out at the ends of the tube, which permits a very natural arrangement for connecting the tube to the input and to the output circuit. At high frequencies, it is usually desirable to bring all ground-return connections to one point, to avoid radio-frequency circulating currents in common impedances formed by shields.

Fig. 13 shows one of the many very useful miscellaneous applications of these tubes. A pentode is mounted at the head end of a shielded flexible cable and used as a probe voltmeter, thus permitting the measurement of voltages at their source.

Conclusions

Because of their decreased lead impedances, interelectrode capacitances and transit times, these miniature tubes allow considerable improvement to be made in high-frequency receiving equipment. It is possible to operate the triode as an oscillator in a conventional circuit down to a wavelength of approximately 40 centimeters. The pentode can be operated as a radio-frequency amplifier down to a wavelength of approximately 70 centimeters. It is practicable to obtain stable gains of

[11] Stuart Ballantine, "Schrot-effect in high-frequency circuits," *Jour. Franklin Inst.*, vol. 206, no. 2, p. 159, (1928).

ten to fifteen at three meters, a wavelength at which the standard tubes are almost entirely ineffectual. Both tubes can be used, down to much lower wavelengths, in exactly the same manner and for the same applications that the corresponding conventional tubes are used; i.e., as oscillators, amplifiers, detectors, converters, and as negative-resistance devices.

The small size of the tubes and their novel structural arrangements allow compact and convenient receiving equipment to be built. Even at the higher wavelengths, they are applicable to a large number of uses for which their size, low weight, and excellent characteristics make them particularly useful.

Acknowledgment

The development of these tubes required a considerable refinement in existing tube manufacturing technique. In this connection, we wish to acknowledge the coöperation of Messrs. S. M. Reed and H. R. Seelen of the Radiotron Developmental Factory. We also wish to acknowledge the contributions of Mr. T. M. Shrader, made during the developmental period of these tubes.

Proceedings of the Institute of Radio Engineers
Volume 23, Number 10 *October, 1935*

A NOTE ON THE SOURCE OF INTERSTELLAR INTERFERENCE*

By

KARL G. JANSKY
(Bell Telephone Laboratories, Inc., New York City)

Summary—*Further consideration of the data obtained during observations on interstellar interference has shown that these radiations are received any time the antenna system is directed towards some part of the Milky Way system, the greatest response being obtained when the antenna points towards the center of the system. This fact leads to the conclusion that the source of these radiations is located in the stars themselves or in the interstellar matter distributed throughout the Milky Way.*

Because of the similarity in the sound produced in the receiver headset, it is suggested that these radiations might be due to the thermal agitation of charged particles.

IN FORMER papers,[1,2] it was explained how interstellar interference was first observed with an automatic field strength recording system making use of a highly directional rotating antenna. It was pointed out that the directions of arrival were fixed in space and that there seemed to be only a single direction of arrival having a right ascension of eighteen hours and a declination of −20 degrees.

Some of the data did not check very accurately the theory of a single direction of arrival and it was suggested that a possible explanation of the discrepancies might be found in refraction of the waves during their passage through the ionized layers of the atmosphere.

Since the publication of the above papers further consideration of the data has led to some very interesting conclusions and speculations. The data obtained from the system are in the form of a continuous record of the output for all hours of the day and, since the antenna rotates continuously, for all directions as well. If we examine a typical day's record of the disturbance, for example that for September 16, 1932, given in Figs. 1 and 2, the following facts are evident. Besides varying gradually in height throughout the day the peaks obtained for each revolution of the antenna also change decidedly in shape. In Fig. 2 from 12:00 M to 3:00 P.M. the peaks are very broad, in fact one peak covers the time taken up by one complete revolution of the

* Decimal classification: R113.5. Original manuscript received by the Institute, April 10, 1935. Presented before Tenth Annual Convention, Detroit, Mich., July 3, 1935.

[1] Karl G. Jansky, "Directional studies of atmospherics at high frequencies," PROC. I.R.E., vol. 20, p. 1920; December, (1932).

[2] Karl G. Jansky, "Electrical disturbances apparently of extraterrestrial origin," PROC. I.R.E., vol. 21, p. 1387; October, (1933).

antenna. From 3:00 P.M. on, the peaks gradually get narrower and narrower until at 10:00 P.M. they are only one quarter the previous

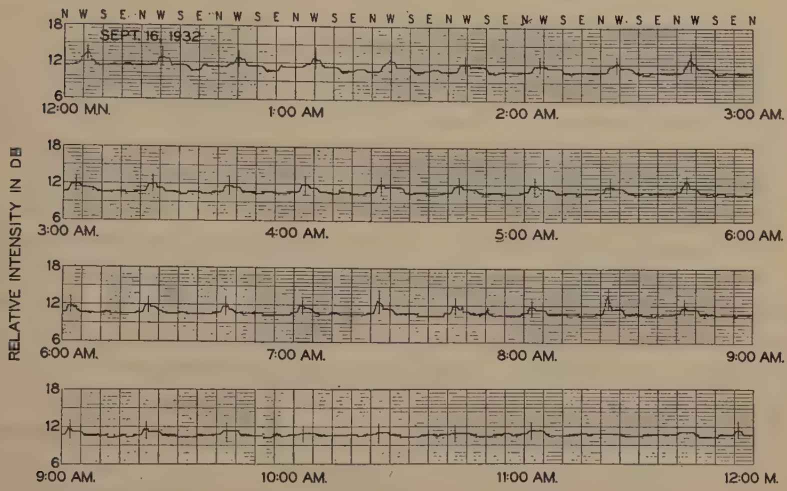


Fig. 1—Sample record of waves of interstellar origin. Record taken 12:00 midnight to 12:00 noon on September 16, 1932.

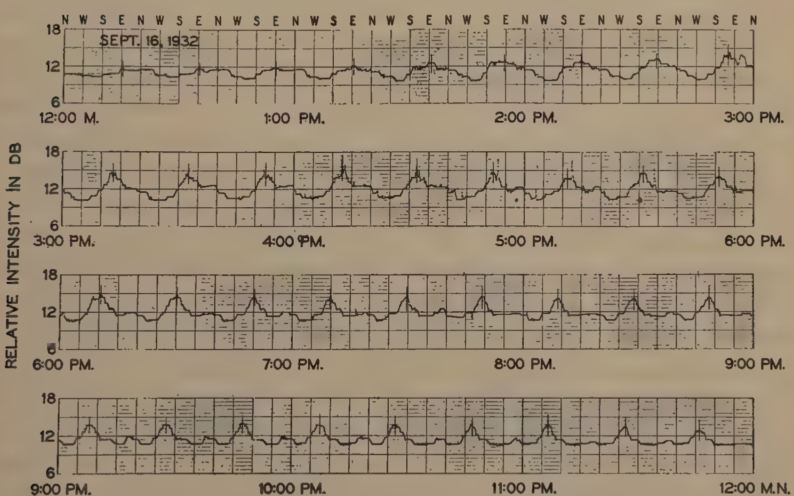


Fig. 2—Sample record of waves of interstellar origin. Record taken 12:00 noon to 12:00 midnight on September 16, 1932.

width. During the same time a much smaller and weaker peak begins to appear on the record. At 9:20 P.M. it is very clear.

Upon determining the direction towards which the antenna sys-

tem points in space at these times, it is discovered that when the peaks are broad, the antenna is so located in space that it sweeps along the Milky Way and the maximum response is obtained when it points in the direction of the center of the Milky Way. When the large sharp peaks and the alternate small peaks are obtained the antenna is so located in space that it sweeps across the Milky Way, the large peak being obtained when that section of the Milky Way nearest the center is crossed and the small peak when that section farthest from the center is crossed.

If we consider the belief now held by astronomers that the Milky Way is a large galaxy of stars having the same general shape as a huge discus or grindstone with the solar system, and therefore the earth, located at some distance from the center and almost in the galactic plane, then the phenomena described above would seem to indicate that the disturbances recorded are due to radiations emanating from the stars themselves. The various heights and widths of the peaks obtained on the record would then be explained in the following manner.

If the axis of rotation of the antenna were perpendicular to the plane of the Milky Way the antenna would rotate so that it always pointed at some part of the Milky Way and therefore would always receive some energy. This energy should reach a maximum value when the antenna points in the direction of the center of the Milky Way System, for the greatest number of stars would then be included within the angle of reception of the antenna. As the antenna rotates the number of stars included within this angle would very gradually decrease until the antenna points in just the opposite direction when the number of stars within the angle would be a minimum. As the antenna rotates further the number of stars within the angle would again increase until the maximum was again reached, etc. Thus the energy received at such a time would show a gradual decrease and increase with one maximum and one minimum for a single rotation of the antenna.

Actually the axis of rotation of the antenna is never exactly perpendicular to the plane of the Milky Way, but approaches that condition closely when the meridian of the receiving location has a right ascension of twelve hours and forty minutes. Translated in terms of the time on the records this occurs about five hours and twenty minutes before the azimuth of the direction of arrival of the disturbances is south. For the record shown in the figures this time would be about 1:50 P.M. at which time the angular distance between the axis of rotation and the perpendicular to the plane of the Milky Way is twelve

degrees and twenty minutes. However, due to the facts that the Milky Way has a very appreciable width, and the vertical directional characteristic of the antenna is very broad, the discussion given above is still applicable and explains the type of record that should be and is obtained at this time. Furthermore, at this time, the center of the Milky Way System as seen from the receiving location has an azimuth very slightly south of east which checks exactly the direction of the maximum disturbance as obtained from the curves.

Since the direction of the axis of rotation of the antenna changes as the earth rotates, the above condition exists for only a short time. Thus, after seven hours and forty-seven and one-third minutes (when the right ascension of the meridian of the receiving location is twenty-two hours and twenty-seven and one-third minutes) the axis of rotation will lie in the plane of the Milky Way instead of being perpendicular to it. In this position the antenna will sweep across the Milky Way twice for every revolution and we would expect two peaks on the record where previously we had only one, a large peak when the antenna sweeps across that part of the Milky Way nearest the center and a smaller one when it crosses that part farthest from the center. These peaks should be relatively sharp because the number of stars within the angle of reception of the antenna changes rapidly. The first case occurs when the antenna points in a southwesterly direction and the second when it points north-northeast.

Turning now to Fig. 2 we see that at 9:37$\frac{1}{3}$ P.M. (seven hours and forty-seven and one-third minutes after 1:50 P.M.) we have two definite peaks on the record for every revolution of the antenna, the larger of which is obtained when the antenna is pointing southwest and the smaller when it is pointing north-northeast, checking the predictions exactly.

After another eight hours and twenty-five and one-third minutes, or at 6:02$\frac{1}{3}$ A.M. on the record given, the axis of rotation of the antenna again lies in the Milky Way, but this time the two points where the antenna sweeps across the Milky Way are both some distance from the center so that neither peak should be very large. At this time the two points have directions of northwest by north and southeast by south, the former being the nearest to the center of the Milky Way. Turning again to the figures we find that at 6:02-$\frac{1}{3}$ A.M. the peaks are much weaker than at 9:37-$\frac{1}{3}$ P.M. and that they have the directions predicted.

A more detailed analysis of the data has shown that every time the antenna points towards some part of the Milky Way the record shows an increase in the energy received, and also every time the record

shows an increase of energy received the antenna is found to be pointing towards some part of the Milky Way.

As said before, the most obvious explanation of these phenomena is one that assumes that the stars themselves are sending out these radiations and that the direction of arrival at the receiving location, instead of being confined to a single direction as was formerly intimated, include all directions, a greater indication being obtained for those directions confined to the Milky Way because of the greater star density there.

Another plausible explanation is one based on an hypothesis previously suggested,[2] that the waves which reach the antenna are secondary radiations caused by some form of bombardment of the atmosphere by high speed particles which are shot off by the stars.

Upon examining the characteristics of these radiations for further clues as to their source, one is immediately struck by the similarity between the sounds they produce in the receiver headset and that produced by the thermal agitation of electric charge.[3] In fact the similarity is so exact that it leads one to speculate as to whether or not the radiations might be caused by some sort of thermal agitation of charged particles. Such particles are found not only in the stars, but also in the very considerable amount of interstellar matter that is distributed throughout the Milky Way, which matter, according to Eddington[4] has an effective temperature of 15,000 degrees centigrade. If the radiations come from such particles one would expect the response obtained to depend upon the directional characteristic and gain of the antenna and the way it is pointed relative to the Milky Way, an expectation which agrees with the observed facts.

Attempting to explain the radiations in question on the basis of any of the above hypotheses immediately raises a serious question as to the effect of the sun. Since the sun is a star, although in comparison with some, a rather insignificant star, and since it is much closer to the earth than the stars of the Milky Way, so close in fact that the energy received from it in the form of light and heat is many times the combined energy received from all the stars and all the interstellar matter of the Milky Way, it would naturally be expected that radiations similar to those in question would be found coming from the sun with an intensity much greater than the intensity of those coming from the other sections of the Milky Way. So far no such solar radia-

[3] F. B. Llewellyn, "A study of noise in vacuum tubes and attached circuits," Proc. I.R.E., vol. 18, p. 243; February, (1930).

[4] Arthur Stanley Eddington, "Stars and Atoms," pp. 66–69, Yale University Press, 1927.

tions have been detected, and, although a possible explanation for their absence might be based on a supposition that the temperature of the sun is such that the ratio of the energy radiated by it on the wavelengths studied to that radiated in the form of light and heat is much less than for some other classes of heavenly bodies found in the Milky Way, the question will have to remain unanswered until more data have been taken.

Proceedings of the Institute of Radio Engineers
Volume 23, Number 10 *October, 1935*

INTERFERING RESPONSES IN SUPERHETERODYNES*

BY
HOWARD K. MORGAN
(Kansas City, Missouri)

Summary—*Superheterodyne receivers respond to other frequencies than the one for which the receiver is tuned. Various types of these undesired responses are treated in particular with one type not generally recognized, the resultant of an off-resonance signal combining with the receiver's oscillator and forming a signal, at the converter tube plate, which lies in the vicinity of the intermediate frequency. High harmonic orders of this are present across the converter plate circuit which recombine, by plate modulation, in the internal plate circuit with the oscillator and form a signal at the intermediate frequency. Methods of reducing this effect are given.*

INTERFERING responses, other than those due to broad selectivity occur in superheterodyne receivers. This is deterimental to services that require the use of receivers in the close proximity of transmitters operating on frequencies in the order of those being received. To distinguish this type from other undesired responses, it is in order to treat briefly the other types as well. Various combinations of the oscillator frequency and its harmonics with incoming signals give undesired responses in the receiver. Almost all interference from this source could be eliminated if the radio-frequency tuned circuits preceding the converter had as much selectivity as the intermediate amplifier. This virtually reduces the superheterodyne to a tuned radio-frequency receiver. Other methods must be devised for reducing unwanted responses to a point where they no longer play an important part.

Four groups of interfering responses will include all frequencies to which the superheterodyne may respond when it is tuned to one desired frequency.

In the following formulas, where subtraction is called for, the smaller frequency should always be subtracted from the larger. The formulas can be written more simply, as has been done, if this is kept in mind.

Group	Fundamental Combination	Harmonic Combination
A	$f = o \pm s$	$f = no \pm ns$
B	$f = o \pm \dfrac{i}{x}$	$f = no \pm \dfrac{i}{x}$

* Decimal classification: R170×R361.2. Original manuscript received by the Institute, October 22, 1934; revised manuscript received by the Institute June 19, 1935.

C	$f = i$	$f = ni$
D	$f = o \pm \dfrac{s}{n}$	$f = no \pm \dfrac{s}{n}$

where,

o = receiver's oscillator frequency
s = frequency to which the receiver is tuned or the image frequency
i = the intermediate frequency
x = the subharmonic order (1,2,3, etc.)
n = the harmonic order (1,2,3, etc.)
f = frequency of interfering signal (and may coincide with s)

Unwanted responses can be found by setting the receiver at a high sensitivity and feeding a high level of (modulated) signal to the receiver from a signal generator and tuning the generator slowly over a wide frequency range centering at the frequency for which the receiver is tuned.

Certain unwanted responses can be located using two signal generators, where one generator supplies an unmodulated signal to the receiver (sufficient to give standard output before modulation is removed from it) and the other generator is swept over a wide frequency range as before and supplies high levels of voltage input. Neither generator is modulated during the test and the output is taken when the whistle which characterizes the unwanted response, usually adjusted for a 400-cycle note, reaches standard output.

In the two-generator test, the frequency of the generator to which the receiver is tuned may be substituted for o in the preceding equations and a second complete set of possible interfering signals may be obtained.

Signal generators should be used which are as free as possible from radio-frequency harmonics. Unidentifiable unwanted responses can often be traced to generator harmonics. By setting the generator to twice the interfering frequency it can be easily determined whether the harmonic was generating the unwanted response as the input necessary for the response will be much lower than when the harmonic was being received. The harmonic may be generated in circuits preceding the converter so that calculations should be made on this basis before deciding definitely that the generator is responsible.

Group A undesired responses will be found irregularly spread over the frequency spectrum. Combinations involving harmonics of the fourth order or greater will be usually too weak to detect. Increased

selectivity ahead of the converter tube will reduce this type of response. Reduced harmonic content generated in the oscillator or radio amplifiers will reduce it as well.

Group B undesired responses are characterized by the fact that they occur near the oscillator frequency and not more than an amount equal to the intermediate frequency on either side of the oscillator frequency. All but two of these are due to integral subharmonics of the intermediate frequency being formed by combinations of the undesired signal and the receiver's oscillator frequencies which are multiplied in frequency in the converter so that they result in an intermediate-frequency signal.

There are two special cases to be considered which are combinations with the fundamental frequency of the oscillator. It will be assumed, to simplify following examples and as it is usually the case where the oscillator condenser is like and ganged with the radio-frequency condensers, that the oscillator is operating at a higher frequency than that to which the receiver is tuned. The equations hold whether this is true or not, but specific examples will not hold. The first case is $f=(o-i)$ which is the frequency to which the receiver should respond, and not an undesired response. The second is $f=(o+i)$ which is the image frequency which is usually the most troublesome undesired response. The image may be reduced by three general methods: additional selectivity before the converter, the selection of a higher intermediate frequency, and phasing methods which tend to reduce the image frequency by voltage cancellation.

Increased selectivity before the converter tube will lower all types of undesired responses in this group.

Group C undesired responses may either originate in the receiver or from outside sources, such as low-frequency transmitters.

Those originating in the receiver are harmonics of the intermediate frequency which have been coupled back to the converter or preceding circuits or the antenna lead-in. Poor intermediate-frequency filtering in the second detector plate circuit may allow strong intermediate-frequency voltages to be passed through the audio amplifier and to the (exposed) loud speaker leads. This may result in energy being fed back to the antenna lead-in. Proper filtering and shielding and placement of parts will cure this type of response which is identified by a heterodyne beat being heard when stations near the intermediate-frequency harmonics are received.

If a signal at the intermediate frequency is fed to the receiver the energy transfer through circuits preceding the converter may be rather high. Intermediate frequencies are chosen by designers where few, if

any, services exist. In order to secure high image ratios the intermediate frequency in future receivers will probably be raised and the ease of selecting a clear channel will vanish. Proper design will make the response from this source reasonably low.

A special condition exists with receivers equipped with an oscillator to heterodyne the intermediate frequency for continuous-wave reception. Harmonics of the oscillator may cause interference near the beat oscillator harmonics but proper filtering and shielding will reduce this sufficiently.

Where the interference comes from the receiver itself an advantage exists since the interference occurs at an unchanging magnitude and may be rather easily eliminated.

Group D undesired responses constitute a class quite different in method of formation from the first three groups. An undesired signal reaching the converter tube is heterodyned, by the receiver's oscillator, and appears at the plate of that tube but differing from the intermediate frequency by an amount up to about forty per cent. The signal is sufficiently attenuated by the selectivity following in the intermediate amplifier so that it is not finally detected. It is obvious that little selectivity is realized in the one tuned intermediate circuit at the converter's plate, so that with a strong undesired signal a high voltage may exist, after combining with the oscillator, on the circuit at some frequency near that of the intermediate frequency. This voltage will become smaller for the same signal input as the resulting frequency becomes further from the intermediate frequency. If a harmonic of this signal at the converter plate combines, due to plate modulation, with the oscillator again and gives the intermediate frequency exactly then there will be an undesired response. As there can be no image selectivity incorporated in a converter tube the harmonic may combine with the oscillator whether it is above or below the oscillator frequency by the same amount. These responses are numerous and many points may be found near to the frequency for which the receiver is tuned. Those giving the intermediate frequency due to subtracting the harmonic frequency from the oscillator are often interleaved with those due to subtracting the oscillator frequency from the harmonic frequency, although different harmonic orders will be involved.

To cite a specific example, which experimental data were secured for, a receiver was tuned to 4500 kilocycles, its intermediate frequency was 500 kilocycles, and the oscillator operated at 5000 kilocycles. Without altering the receiver tuning, a strong 4550-kilocycle signal was fed to the receiver. It combined with the oscillator to give 450 kilocycles at the converter tube plate. The tenth harmonic of 450 kilocycles which

is 4500 kilocycles recombined by plate modulation with the oscillator and gave the intermediate frequency and a response.

On first analysis the seriousness cf this type of undesired response might be doubted. The 4550-kilocycle signal was only attenuated to one fifth its value in the radio-frequency circuits. The 4500-kilocycle harmonic due to inefficiencies involved in the harmonic generation, was attenuated to one two-thousandths of the 450-kilocycle value. The undesired response then appeared with a total attenuation of ten thousand times in comparison to the attenuation of the frequency for which the receiver was tuned. While this ratio is fairly high it is insufficient for receivers operating near transmitters which might be operated on the 4550-kilocycle frequency.

Nothing can be done in the grid circuit of the converter tube as the combination occurs in the tube. This point was verified experimentally. A higher Q coil in the plate circuit of the detector will reduce some of the responses materially but generally little additional attenuation will be secured.

The amount of gain before the converter tube has but a small effect as well.

There are but two methods of reducing this type of response.

The first is to reduce the amount of excitation from the oscillator to the converter. This will improve the ratio because the oscillator enters twice into the production of the undesired response while it enters but once for the desired response. Thus if the oscillator excitation voltage is reduced to a tenth of its original value and as the interfering signal combines twice with the oscillator it will be reduced to one hundredth of its previous value. The desired signal will be reduced to a tenth of its previous value. A net improvement in ratio of ten is secured and in the example given the undesired response will be one one-hundred-thousandth of the value of the desired signal. With proper gain before the converter, it has been found experimentally, that the receiver's signal-to-noise ratio is not altered. Thus no disadvantage exists other than that more gain is required in the receiver to make up for converter efficiency loss.

Oscillator coupling to the converter cannot be reduced in the conventional pentagrid converter circuits as the electron stream is fully modulated by the oscillator section. If the pentagrid is coupled to a separate oscillator by utilizing its oscillator grid to feed the energy to the converter section the excitation can then be easily controlled, and several advantages result in this method of coupling for oscillator stability, ease of automatic volume control for the converter, etc.

Another scheme for the reduction of the group D responses is in

the use of a push-pull converter circuit utilizing two converter tubes. While all other parts of the tube are operated as a push-pull circuit the control grids are tied in parallel and are fed from the tuned radio-frequency circuit. Thus the desired signal is heard, as the oscillator excitation on the two tubes is 180 degrees out of phase but the second combinations (group D responses) are not in proper phase for addition and will cancel to a marked degree. Group C interferences will be markedly reduced as well, as straight-through amplification is balanced out with this type of circuit.

An orderly method of rapidly spotting the groups to which interferences probably belong will be the saving of laboratory time.

One of the first things to be noted when listening to the undesired response is the angular amount of turning the signal generator dial over which a signal can be heard. In initially adjusting the receiver, in the example given, the generator at the desired frequency was adjusted to give about 50 milliwatts of audio output with a small signal fed to the receiver. The signal could be heard over ten degrees of motion of the generator dial and beyond this the signal dropped to inaudibility. When the input from the generator was raised with the generator at the undesired response frequency so that 50 milliwatts of signal could be heard, it was found that the signal could be heard but over one degree of signal generator dial motion. Apparently the selectivity of the receiver was changed but in reality the tenth harmonic moves ten times as fast in frequency as the fundamental and this is responsible for the difference noted. With practice one is able to judge the angular motion in passing the generator through the undesired frequency and estimating the probable harmonic involved.

With this principle in mind the harmonic order involved for any type of undesired response may be found and this is particularly useful for group D types.

Group A undesired responses fall at dissociated points. Calculation of the eight frequencies involving the fundamental and second harmonic will usually be the quickest method of identifying these.

Group B undesired responses can be quickly identified as they exist on both sides of the oscillator at one half, one third, one fourth, etc., of the intermediate frequency in difference from the oscillator frequency. The image exists at the intermediate-frequency difference on one side of the oscillator and the desired signal at the same distance on the other side.

Group C responses can be found by tuning the receiver through points at harmonics of the intermediate frequency which fall within the tuning range or through harmonic points of the intermediate beat

oscillator where one is used. A signal should be fed to the receiver at the same frequency for which the receiver is tuned in making this test. A whistle will identify the interference which is coming from the receiver itself.

For external interference tests a generator should be connected to the antenna circuit and a high level of intermediate frequency, or its subharmonics, should be fed to the receiver. With this signal unchanged the receiver should be tuned over its full range to determine the points of worst response.

For group D the "apparent sharpness of signal" method outlined may be employed with a modulated generator set at high inputs slowly tuned over a wide frequency range. The points will generally be found rather close to the receiver's tuned frequency. Points falling within ten kilocycles of the frequency for which the receiver is tuned may be better identified by using one generator without modulation and listening for a whistle. Early stages in the receiver may be overloaded in doing this and it is better to move the receiver tuning a few kilocycles and repeating the first part of the test. Undesired responses will change in relation to the signal if they were previously masked by overloading and may be identified after slightly altering the receiver tuning. Usually if one of the group D interferences is bad, they all will be of about the same order so that it is unnecessary to be concerned over masked interferences until final tests are made.

In all receivers which have been tested in actual operation near an operating transmitter, those with small excitation coupling to the converter are outstandingly better in lack of undesired responses This verifies the laboratory results. By field test the group D interferences have been found to be serious in magnitude in receivers employing high excitation for the converter. These responses should be reduced to the neighborhood of several hundred thousand times for satisfactory operation near transmitters.

Proceedings of the Institute of Radio Engineers
Volume 23, Number 10 *October, 1935*

THE PRESENT STATE IN THE ART OF BLIND LANDING OF AIRPLANES USING ULTRA-SHORT WAVES IN EUROPE*

By
E. Kramar
(C. Lorenz Aktiengesellschaft, Berlin-Tempelhof, Germany)

Summary—*In a previous issue of the* Proceedings[1] *we published a paper concerning the application of ultra-short waves in connection with navigation, based on the glide path proposed by Diamond and Dunmore. That paper not only explained the advantages of using radio beacons for this range of waves but also made mention of the fact that it is deemed expedient to use radio beacons operating on ultra-short waves for blind landing of airplanes. Meanwhile this idea has been developed further so that we are now in a position to report on the results of tests made with this method of blind landing. At present this method is on the verge of being introduced in Europe on a large scale. Besides the beacons installed in Berlin and Zürich which have been in service for a considerable length of time, the German airports at Hannover, Cologne, and Frankfort (Main) will be equipped with radio beacons by the time this paper is published. Similar equipments are in course of installation in the German airports of Munich and Königsberg, and in the coming months will be installed in France, England, Poland, Austria, Sweden, Czechoslovakia, Denmark, and Japan.*

REGULARITY and reliability of service, as a matter of principle, are the fundamental requirements for any conveying means including aircraft. In view of the exceedingly high speed of aircraft the aviation technique above all demands a very high degree of refinement in design for the navigating instruments, the use of which becomes of utmost necessity when there is no visibility to ground and when the pilot cannot check the altitude of his airplane on the one hand and the direction of movement as well as the respective position on the other except by means of his navigating equipment.

Whereas the instruments for use in blind flying have for a considerable time been advanced to such an extent that with proper training the pilot has no difficulty in flying within or above the ceiling, wireless electrical methods which will allow aircraft to approach an airport and then land with perfect safety in case of poor or zero visibility have been but recently developed.

I. Fundamentals of Operation

For international European aviation it was found necessary to develop the means for blind landing in such a way as to allow of its being easily added to the existing ground radio stations. In Germany a trial

* Decimal classification: R528. Original manuscript received by the Institute, April 29, 1935.
[1] E. Kramar, "A new field of application for ultra-short waves," Proc. I.R.E., vol. 21, pp. 1519–1532; November, (1933).

plant for bad weather landing, manufactured by the C. Lorenz A. G., was installed in the airport of Berlin-Tempelhof in 1932 under the auspices of the German Ministry of Aviation. This system was demonstrated to the members of the International Aviation Conference in Berlin in January, 1933. The subsequent conference (Geneva, 1934) as well as the meetings held by experts on aviation wireless matters (Paris, November, 1933, and Warsaw, September, 1934) dealt with the state of this system and fixed the wave for the landing beacon (9 meters, 33.3 megacycles) as well as that for the marker beacons or entrance signals (one wave between 43.0 and 33.3 megacycles) and also determined the modulating frequencies for the individual transmitters. These resolutions and the tests meanwhile carried through with the system have led to starting the installation of a landing radio beacon system to serve Germany as well as part of the neighboring countries.

II. Description of the Ground Station

The following is a detailed description of the technical features of the system which has been built by the C. Lorenz A.G. in accordance with the specifications of the German Ministry of Aviation. The tasks to be fulfilled by the individual transmitters are as follows:

1. Determination of the direction of approach; i.e., of the sector of approach not endangered by obstacles, a range of about 30 kilometers (altitude 400 meters) being demanded, corresponding to a time of approach of about 8 to 10 minutes.

2. Transmission of the entrance signals; i.e., signals marking the distance from the airport, the transmitter of the first or warning signal preferably to be located where the pilot is to start landing.

3. Vertical navigation; which problem will best be solved by the use of a glide path as proposed by Diamond and Dunmore[2], to assist the pilot when his airplane is gliding down. The glide path begins at the warning signal and ends, in the case of small landing places, at the main signal and in the case of large places at a certain point of landing in the airport.

Problems 1 and 3 are solved by means of the same ultra-short wave (9 meters, 33.3 megacycles), in the case of the second problem a wave of 7.9 meters (38.0 megacycles) is used. This wave band was chosen because it offers the following advantages:

Limited range, thereby rendering it possible to have all airports operate at the same wavelength, without interfering with one an-

[2] H. Diamond and F. W. Dunmore, "A radio beacon and receiving system for blind landing of aircraft," *Bur. Stan. Jour. Res.*, vol. 5, p. 897, (1930); Proc. I.R.E., vol. 19, p. 585–627; April, (1931).

other, the receiving set requiring no adjustment in going from one airport to another making possible automatic operation.

Discrimination between communication service on medium waves and navigation for landing on ultra-short waves so as to eliminate interference between said circuits.

No atmospheric disturbances and no night effects in this wave band.

Simple and safely operating beam antenna systems of highest efficiency of radiation and of small dimensions at the transmitting end.

Possibility of combining the approach and the glide path on the same wave.

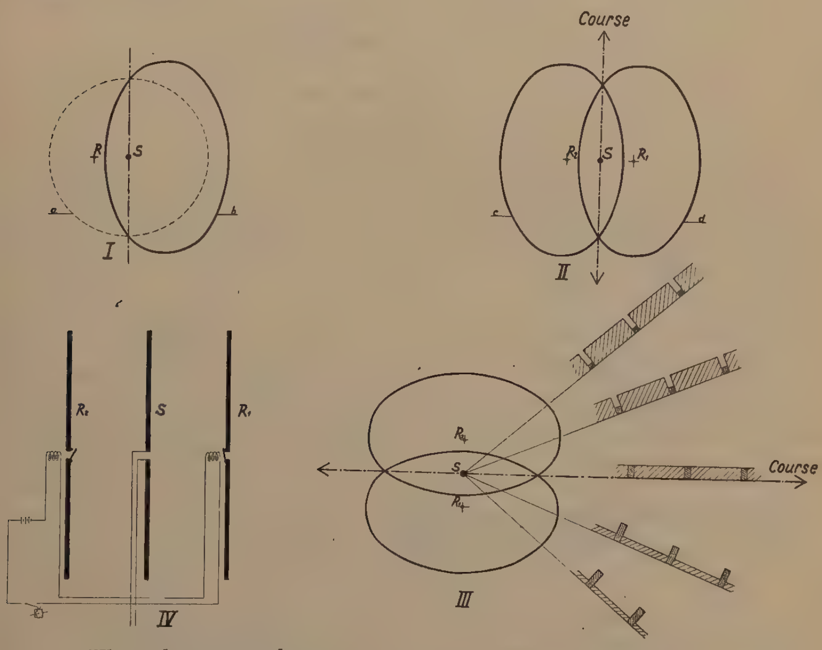


Fig. 1—Ultra-short-wave beacon.
Propagation diagrams
I. (a) Transmitting dipole S alone.
(b) Transmitting dipole S with reflector *R*.
II. (c) Transmitting dipole S with reflector R_1.
(d) Transmitting dipole S with reflector R_2.
III. Formation of guide ray.
IV. Principle of reflector keying.

The direction of approach is indicated by a guide ray beacon of the reflector type.[3] Independent of position, altitude, course of the air-

[3] British Patent, 405,727, C. Lorenz A.G., Berlin; Prior, Germany, September 2, 1932; German Patent, 616,026, C. Lorenz A. G., Berlin, April 6, 1934.

plane, and wind drift the pilot is in a position to know from the indications received from the ground whether he is approaching the airport from the proper direction. The method of transmission has already been reported on previously in the PROCEEDINGS, so that Fig. 1 may suffice as a reminder.

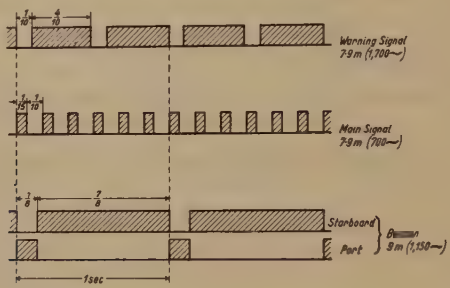


Fig. 2—Keying and modulation characteristics of entrance and beacon signals.

Each of the two distance or entrance signals is provided by a small transmitter feeding a horizontal dipole. With the proper distance from ground, a "wall" of signals is brought about which the airplane has to pass when approaching. The two signals differ in modulation and keying (Fig. 2).

With respect to the glide path, viz., vertical navigation, the idea of the airplane gliding down on a line of equal field strength is made use of as developed by Diamond and Dunmore. However, compared with the latter method, the system now described shows two principal differences:

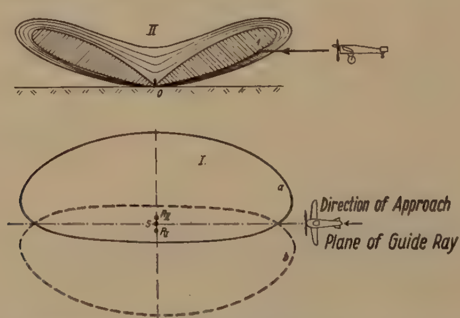


Fig. 3—Vertical and horizontal directional characteristics of beacon transmitter.

1. In order to produce this glide path the same transmission arrangement is used as for determining the direction of approach; i.e.,

vertically polarized radiation from a vertical antenna is employed.[4] (Fig. 3.) This provides a field strength pattern similar to that produced by the horizontally polarized antenna array developed by Diamond. This saves not only a special transmitter with its directional system but above all saves a special receiver on board the airplane.[5]

2. It is possible to calibrate the receiver by noting the beacon signal indication when intercepting the landing beam at the moment the warning signal is received, the plane being at a known altitude of approach (Fig. 4). If, for instance, a certain altitude of approach is fixed

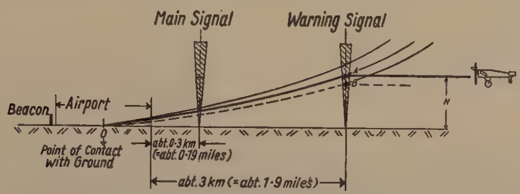


Fig. 4—Diagram illustrating method of utilizing simultaneously the entrance and beacon signals.

which the altimeter indicates at a precision of about ±10 per cent, the plane of the warning signals cuts out, from the multitude of field strength lines, the one which independent of its absolute magnitude characterizes the glide path. Independent of the transmitter output and of the sensitivity of the receiver it is, therefore, only necessary when gliding down to adhere to this particular intensity of beacon signal; i.e., the deflection indicated by the respective instrument when the airplane passes the warning signal. Only during the few minutes of landing does the radiation of the transmitter and the sensitivity of the receiver have to be constant and this may readily be brought about. The absolute measurement of the field strength is thus changed into a relative measurement which considerably increases safety of operation.[6]

Some technical data for the system are given below:

Fig. 5 shows an airport where one landing beacon and four transmitters of entrance signals are provided; the whole system is remote controlled from the direction finding station communicating with the airplanes. Depending upon the wind direction, one or the other sector may be beacon-marked, the glide path to be used, however, only from one side, inasmuch as where this is employed the airplane is allowed

[4] French add. Patent, 44,879, C. Lorenz A. G., Prior, Germany, July 24, 1933.

[5] British Patent, 408,321, C. Lorenz A.G., Berlin; Prior, Germany, January 6, 1932.

[6] German Patent, 607,237, C. Lorenz A.G., Berlin, Feburary 22, 1933.

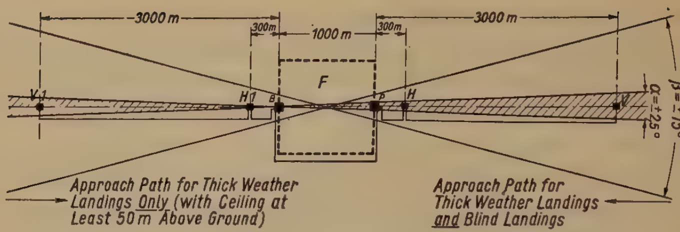


Fig. 5—Diagram of arrangement of radio beacons operating on ultra-short waves for thick weather landings and blind landings at commercial airports in Germany.

F	Airport	V, V_1	Warning signal
----	Boundary of airport	α	Guide ray
P	Direction finding station	β	Sector free of obstacles
B	Approach beacon	—·—	Center line of landing path
H_1H_1	Main signal	——	Remote control cable

Fig. 6—Beacon transmitter.

only to approach, but not to pass, the beacon. The remote control makes it possible to operate the beacon by one pair of signals or the other; accordingly the keying of the beacon is automatically changed

over in such a way that, always looking in the direction of approach, dots are on port and dashes are on starboard.[7] The idea of this method is to relieve the pilot of unnecessary brain work and to save him from switching the direction indicator, placing all the responsibility for the proper approach on the ground station.

The beacon transmitter, Fig. 6, is quartz controlled, operating at a frequency of 33.3 megacycles; it has five stages with an antenna out-

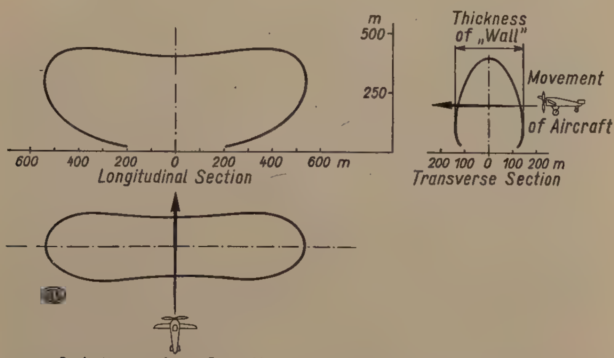


Fig. 7—Radiation diagram of signals (according to measurement flights).

put of 500 watts and 90 per cent modulation with a frequency of 1150 cycles. It is fed solely by selenium dry rectifiers. The maxinum height of the antenna system is seven meters.

The main and warning signal transmitters (38 megacycles) are likewise quartz controlled, three stages, providing an antenna output of 5 watts and 90 per cent modulation (700 and 1700 cycles). The horizontal dipoles are arranged about four meters above a specially shaped reflector surface and are made so as to give a diagram as shown in Fig. 7.

The remote control apparatus (Fig. 8) is arranged at the direction finding station and serves for cutting in and out the beacon and the associated groups of signals. This instrument also allows of a satisfactory supervision of the operation of the whole system; in case of trouble it is possible to locate the fault (cable, power supply, or transmitter) at once, so that the trouble may be eliminated promptly.

III. Equipment on Board the Aircraft

In designing the receiving set the features stated below had to be taken into account: Full automatic operation; i.e., no attention re-

[7] German Patent, 612,825, C. Lorenz A. G., Berlin, May 8, 1934.

quired during the approach; highest degree of safety; i.e., limited number of valves, low weight.

The receiver of the beacon of approach (wave of nine meters modulated with 1150 cycles) is provided with a high-frequency stage of amplification, a detector, and two low-frequency amplifying stages. Automatic control of amplitudes takes care of converting the excessively large variations of the field intensity during the approach into

Fig. 8—Remote control of apparatus.

allowable variations of volume in the head receiver.[8] The volume of beacon signal is indicated by an instrument of the rectifier type (indication of distance and glide path). Furthermore the special instrument for indicating the lateral deviations (port and starboard deviation) is connected to the second low-frequency stage through a rectifying tube (Fig. 9).[9]

[8] British Patent, 424,275, C. Lorenz A. G., Berlin; Prior, Germany, December 12, 1932.

[9] French Patent, 786,613, C. Lorenz A. G., Prior, Germany, August 14, 1934.

A second detector is provided for the reception of the warning and main signals; the low-frequency amplifier of the nine-meter receiver is also used for amplifying the modulating frequencies of these signals.

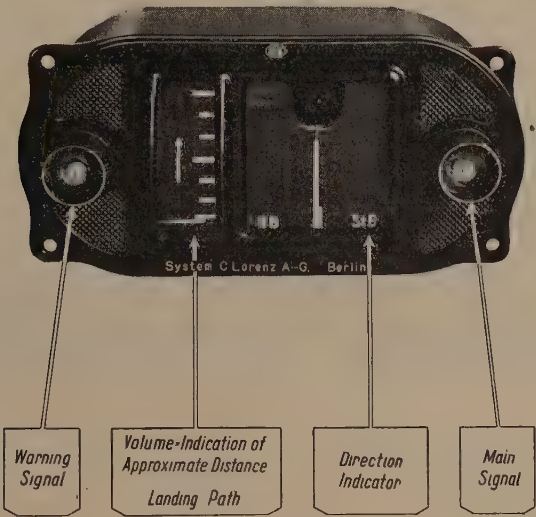


Fig. 9—Visual indicator assembly.

A head receiver is directly connected to the second low-frequency stage so that all of the three frequencies may be received acoustically. For the visual indication, however, the frequencies must be separated; this is done by a special band-pass filter. As may be seen from Fig. 10, the two instruments respond only to the beacon frequency both as re-

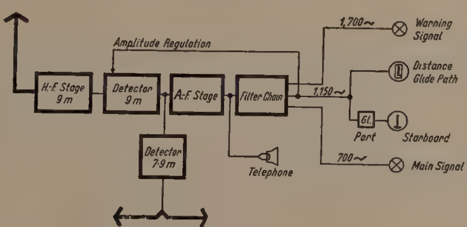


Fig. 10—Diagram of receiving circuits.

gards indication of distance and glide path on the one hand and of lateral deviations on the other hand. The same holds true for the automatic amplitude control circuits which control only the nine-meter channel. The warning and main signals thus do not affect the two indicating instruments or the amplification of nine-meter signals. Each one of the two entrance signal modulations causes a small glow lamp to flicker in the rhythm of keying the signals. This arrangement is neces-

sary in order to eliminate as far as possible any mistakes concerning the signals as well as erroneous indication.

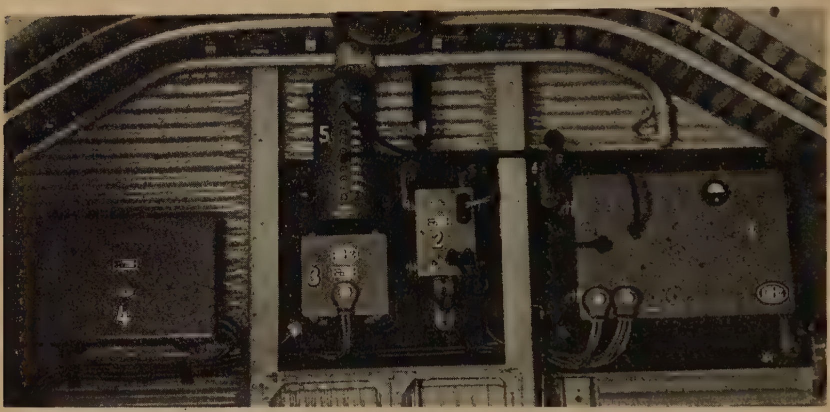


Fig. 11—Installation of the receiver. (Aluminium coverings of airplane removed.)
1. 9-meter receiver
2. 7.9-meter detector
3. Filter chain
4. Battery box
5. Loading coil of long-wave transmitter (the latter does not belong to the ultra-short wave equipment).

Fig. 12—Visual indicator assembly on the pilot's instrument panel.

A vertical rod of about 80 centimeters in length serves as aerial for the nine-meter wave in the airplane. The entrance signals require a horizontal dipole; it consists of two wires of 80 centimeters in length,

arranged in longitudinal direction underneath the body of the airplane. A spacing of about five centimeters from the metal surface is considered sufficient so that the air resistance is negligible.

Inasmuch as the receiving equipment requires no attendance or adjustment whatever, it need not be accessible and is, therefore easy to mount in an airplane. Fig. 11 shows the receiving set mounted in a "Junker 52" cargo aircraft of the Deutsche Lufthansa; Fig. 12 shows the instrument mounted on the pilot's instrument panel.

IV. Landing Performance Using the System Described in Sections II and III

It is assumed that by some sort of straight-line direction finding the airplane has come within the range of the landing beacon; i.e., for about thirty kilometers round the airport. From this moment when the pilot visually from the instruments or aurally by means of his head receiver perceives the signals sent by the beacon—the way of reception is left to the pilot's choice, both indications being given simultaneously—he will try to come as quickly as possible in the correct direction of approach which is prescribed by the continuous dash zone of the beacon. In his further approach the two instruments enclosed in a common casing advise him on the one hand of his position with respect to the proper direction of approach; i.e., port or starboard deviation, respectively, and on the other hand advise him of his approaching the airport by the increase in volume of beacon signal. The deflection of the needle to the latter effect gives a rough measurement of the distance. The pilot will now slowly go down to an altitude of 200 meters and will wait for the warning signal. At the moment the pilot hears the characteristic signal and perceives the flickering of one of the glow lamps he switches his receiving set to "glide path" and throttles the motors; when gliding down he tries to adhere to, or to reach again, until he arrives at the main signal, the particular deflection now indicated by the beacon signal measuring device; flying below this mark means danger. The main signal, discernible by the deep sound on the flickering of the other glow lamp, indicates that the airplane is near the boundary of the airport.

The pilot when gliding down is also advised of the exact direction visually and aurally in the same manner as during the preceding approach.

The system described in this paper has been thoroughly tested under the auspices of the German Ministry of Aviation in coöperation with the Deutsche Lufthansa and the Deutsche Versuchsanstalt für Luftfahrt (German Research Institute for Aviation). More than 1000

training flights made by the Deutsche Lufthansa offered a good chance for studying in detail the operating requirements in order to be able to comply as far as possible with the requirements of actual service. Already the Deutsche Lufthansa has equipped about twenty per cent of all aircraft with the necessary equipment and plans to provide all aircraft of the most important air routes with landing receiving sets by next fall. By that time all of the larger airports in Germany will have been equipped with the corresponding ground stations.

Proceedings of the Institute of Radio Engineers
Volume 23, Number 10 *October, 1935*

AN ANALYSIS OF CONTINUOUS RECORDS OF FIELD INTENSITY AT BROADCAST FREQUENCIES*

By

K. A. Norton, S. S. Kirby, and G. H. Lester
(National Bureau of Standards, Washington, D.C.)

Summary—*Continuous records of the field intensities of most of the broadcast stations in the United States have been made at the National Bureau of Standards receiving station near Washington, D. C. Typical records of received field intensities from several stations are presented. Maximum field intensities during ten-minute time intervals are analyzed in the following ways to illustrate sky-wave propagation phenomena at broadcast frequencies for distances up to 4000 kilometers. (1) The diurnal variation of the ten-minute maxima is given for several stations. (2a) The variation of the ten-minute maxima is shown with respect to distance for night field intensities. (2b) These variations are also shown for sky waves received during the daytime. Empirical formulas are developed to represent the data of (2a) and (2b). The data are explained in terms of a theory of propagation of waves in the ionosphere.*

During the past three years graphical records of the field intensity of over three-hundred broadcast stations in the United States and Territories have been made at the National Bureau of Standards receiving station at Meadows, Maryland, near Washington, D.C. The method used for recording the field intensity is described elsewhere.[1] It is the purpose of this paper to report these data and some of the conclusions reached from an analysis of these records.

I. Experimental Data

IN ORDER to illustrate the type of field intensity record obtained and some of the general characteristics of the phenomena under consideration, some typical records are given of the received field intensity from broadcast stations WLW in Cincinnati, Ohio, and WCKY in Covington, Kentucky. The upper record in Fig. 1 shows the received field intensity of WLW as measured at the Meadows field station (latitude 38° 48′ 32″N., longitude 76° 52′ 40″W.). From 6 A.M. to 1 A.M. the power used was 50 kilowatts and from 1 A.M. to 5:30 A.M. the power used was 500 kilowatts. It is evident from the records that transmissions were stopped intermittently between 2 A.M. and 6 A.M. The inverse distance values were calculated on the assump-

* Decimal classification: R270. Original manuscript received by the Institute, October 29, 1934. Presented before Ninth Annual Convention, Philadelphia, Pa., May 30, 1934, and before General Assembly, International Scientific Radio Union, London, England, September 12–18, 1934. Published in *Bur. Stan. Jour. Res.*, vol. 13, pp. 897–910; December, 1934. Publication approved by the Director of the National Bureau of Standards of the U. S. Department of Commerce.

[1] K. A. Norton and S. E. Reymer, "A continous recorder of radio field intensities," *Bur. Stan. Jour. Res.*, vol. 11, pp. 373–378; September 1933, (RP 597).

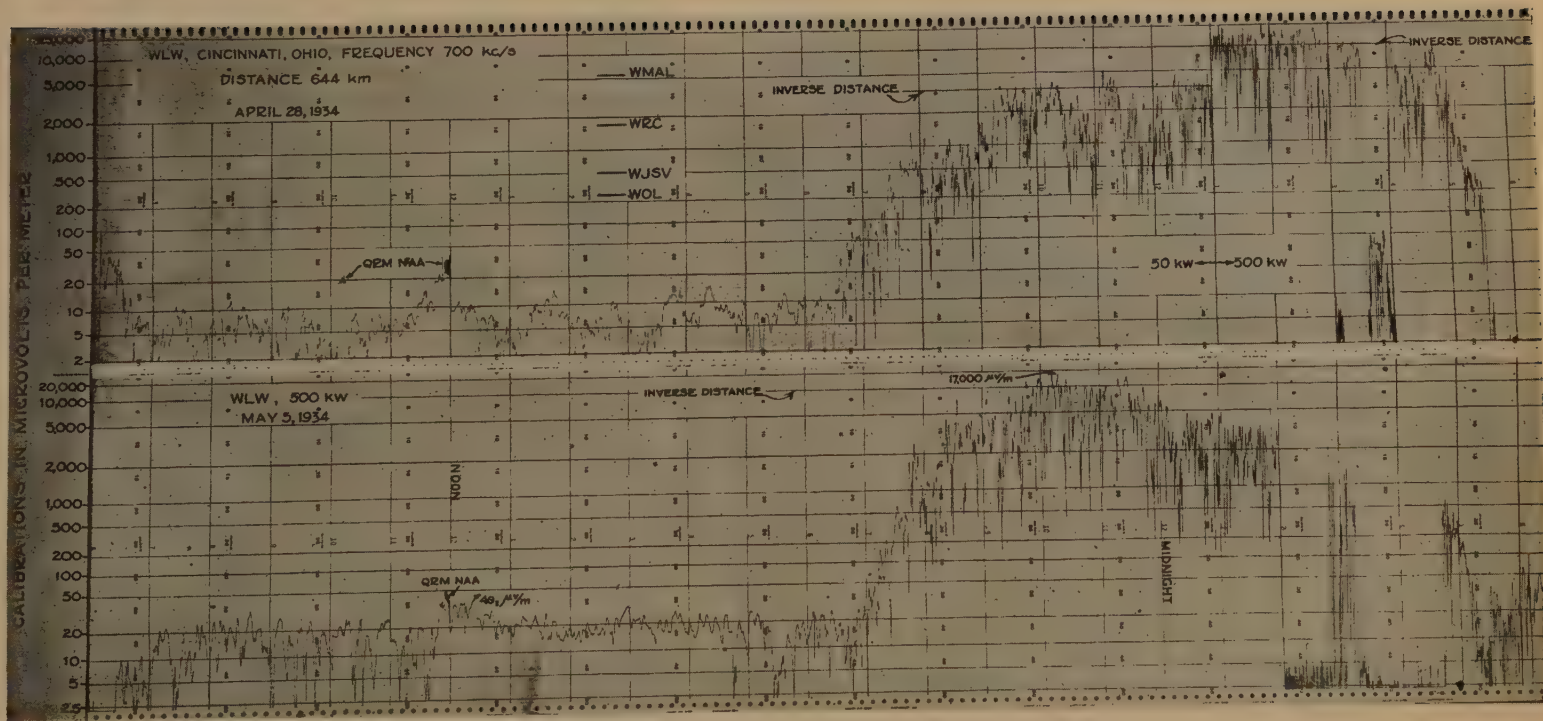


Fig. 1—Graphical field intensity record of WLW, Cincinnati, Ohio.

tion that these indicated amounts of power were actually radiated. The maximum received field intensities reach values two and one-half times the inverse distance field intensities. The fading observed throughout the daytime indicates a received sky wave of intensity comparable to that of the ground wave since ground-wave field intensities are free from fading. In order to indicate the relative intensity of this station the intensities of the ground waves of the four local Washington stations are shown: WMAL, 630 kilocycles, 500 watts, 14 kilometers; WRC, 950 kilocycles, 500 watts, 19 kilometers; WJSV, 1460 kilocycles, 10 kilowatts, 15 kilometers behind directional antenna; WOL, 1310 kilocycles, 100 watts, 17 kilometers. On the lower record 500 kilowatts were used throughout the day. The peak field intensity was only 17 millivolts per meter on this day as compared to 25 millivolts per meter a week earlier. Such a variation is typical of night field intensities at broadcast frequencies. The period of the night fading at this frequency is about five minutes.

Fig. 2 shows two records of WCKY, the upper record for June 23 near the summer solstice and the lower record for December 20 near the winter solstice. The most conspicuous difference between the two records is the absence of any recordable field intensity during the summer day and the relatively strong sky wave present during the winter day. The peaks which may be seen during the summer day represent interference from other stations, only a weak beat note being audible for WCKY. The noise level at night is about eight times as strong in the summer. Also the night field intensities are about twice as strong in the winter record. Notice particularly the night fading at this frequency which has a period of about one and one-half minutes. The record from midnight to 7 A.M. represents atmospherics.

In order to condense the enormous amount of data obtained, our records were analyzed in terms of the peak field intensities which were observed during ten-minute intervals. It may be mentioned that the average of six consecutive ten-minute maxima corresponds closely to the quasi-maximum field intensity for the hour; i.e., the field intensity which is exceeded during the hour only five per cent of the time. In order to make some correction for the power it was assumed that each station radiated one half of its rated power. Admittedly this was a very poor assumption but it was about the best that could be done in view of the fact that for most stations the amount of power which is radiated is not known. As an illustration of how poor this assumption regarding radiated power may be in some cases it may be mentioned that WLW with a rated power of 500 kilowatts actually radiates at low angles power equivalent to 870 kilowatts for the ordinary vertical dis-

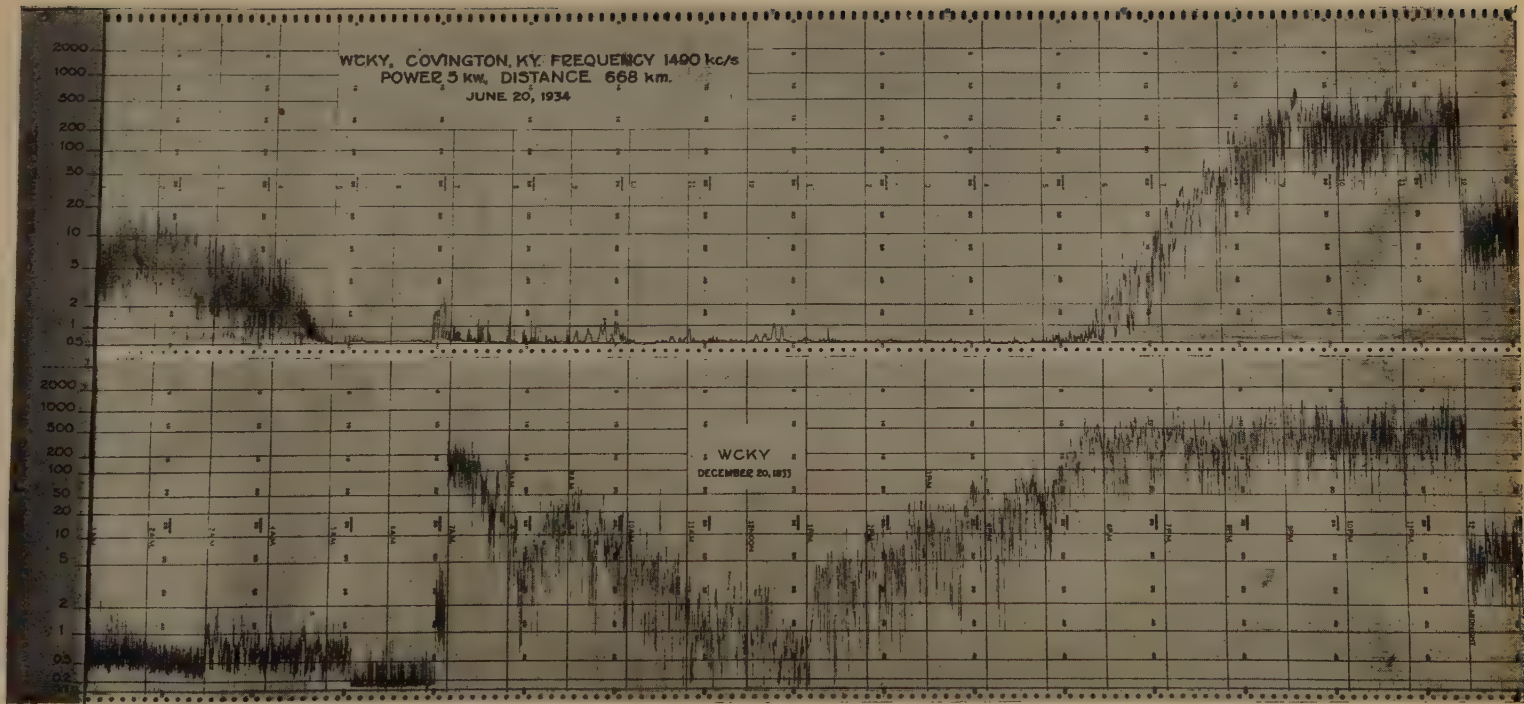


Fig. 2—Graphical field intensity record of WCKY, Covington, Kentucky.

tribution assumed for other stations. The usual formula for the radiated power for a vertical infinitesimal doublet is used in these calculations; i.e., $P_r = F^2D^2/C^2$ kilowatts, where F is the measured ground-wave field intensity in microvolts per meter at a distance D in kilometers, and C is the velocity of light in kilometers per second. In the few cases where field intensity measurements have been made near the station, the radiated power was calculated by means of the above formula in terms of these measurements. Such cases were the exception rather than the rule and still do not take into account the radiation characteristics in the vertical and horizontal planes.

Fig. 3 shows the diurnal variation of the ten-minute maxima for several clear-channel broadcast stations. They are shown on a logarithmic scale with the calculated inverse distance field intensity equal to one. If we write $F = AC\sqrt{P_r}/D$, then each measurement of field intensity determines an attenuation factor, A; these are the values plotted in Fig. 3. The principal thing to be noticed here is the increase in the ratio of the night field intensity to the day field intensity with distance. The increase in this ratio is evident in spite of the fact that it was impossible to keep all of the variables such as frequency and terrain constant. It may be seen that these attenuation factors are greater than one for all the stations at some time during the evening, the highest peaks occurring for WGN which reaches a peak of four, corresponding to four times the calculated inverse distance field intensity. Of interest also is the fading observed at the greater distances during the daytime indicating the presence of a sky wave of intensity at least comparable to that of the ground wave. Evidence will now be given to show that this sky wave during the daytime is much stronger than the ground wave for distances greater than about 600 kilometers.

Fig. 4 shows the received daytime field intensities for a number of stations corrected to one kilowatt of radiated power as a function of the distance at those distances for which there is fading throughout the day. Each point represents the average of the ten-minute peaks received between 10 A.M. and 2 P.M. The open circles represent our measurements made in December, 1933, the crosses are for January, 1934, and the solid circles are for April, 1934. The solid lines indicate the theoretical variation of the received field intensity of the ground wave at 1000 kilocycles for 10^{-13} and for 10^{-14} electromagnetic units conductivity as taken from the Report of the U. S. Committee on Radio Propagation Data.[2] The boxes represent measurements made on WLW by their engineers. The open boxes are for measurements east of WLW and the boxes with crosses are for measurements north-

[2] Proc. I.R.E., vol. 21, pp.1419–1439; October, (1933).

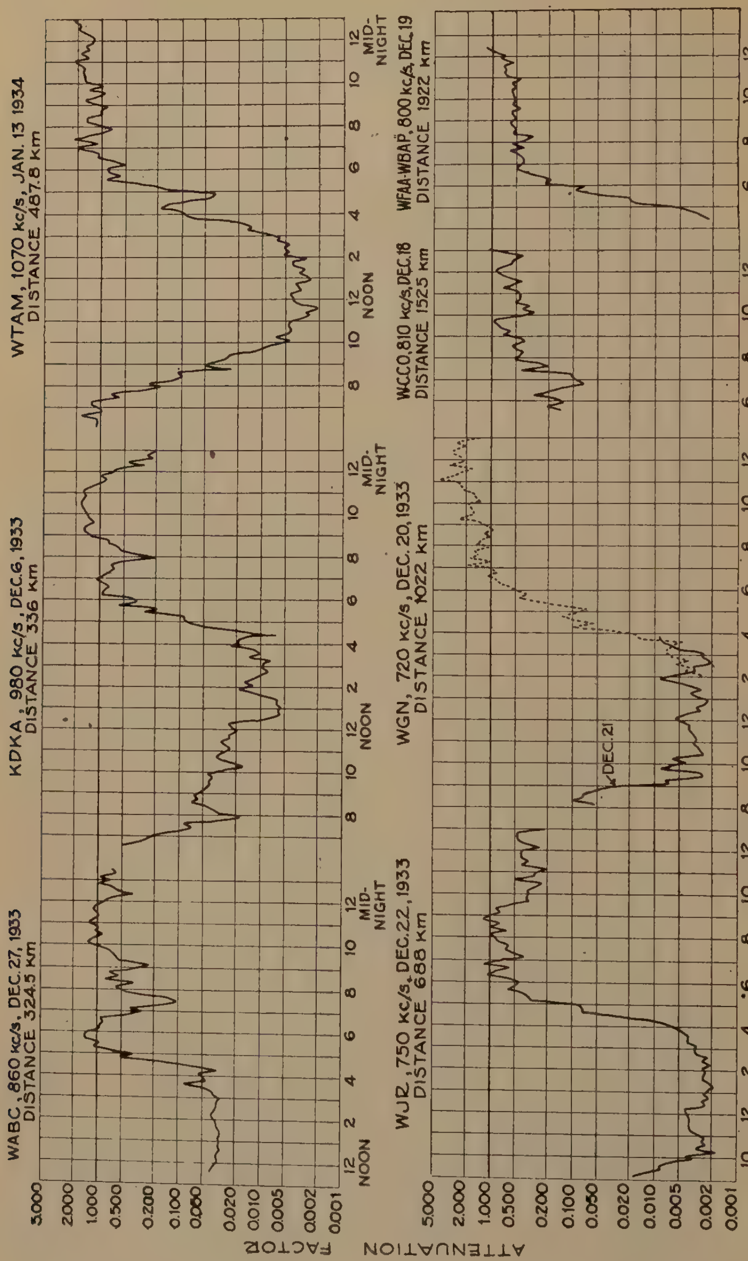


Fig. 3—Sample diurnal variations of the attenuation factors of several broadcast stations given as a function of distance.

east of WLW. All of the measurements of WLW were made in February and March, 1934, when snow was on the ground. It is probable that the received field intensities at distances greater than about 600 kilometers are primarily due to sky waves since the variation with distance departs from the theoretical ground wave curves at these distances.

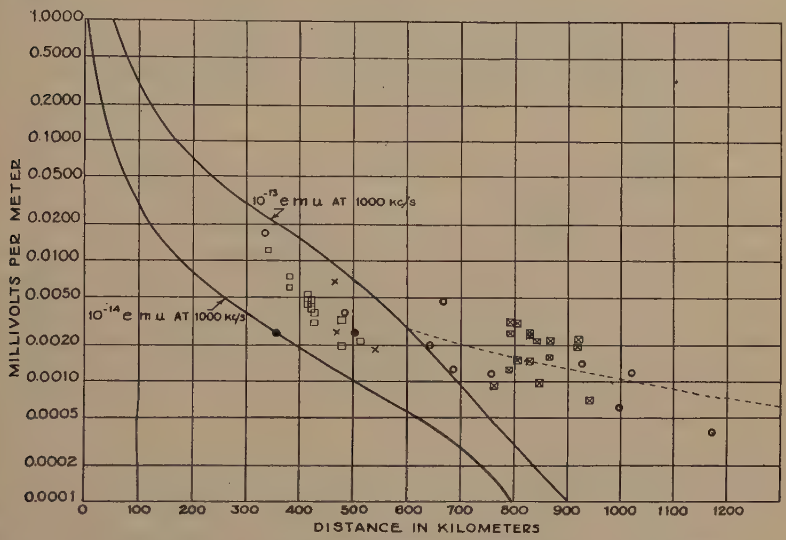


Fig. 4—Daytime field intensity measurements reduced to one kilowatt of radiated power and given as a function of distance for broadcast stations in the United States. The solid curves were taken from the Report of the United States Committee on Radio Propagation Data and are theoretical ground-wave formulas. The dotted curve represents an empirical formula for distances greater than 600 kilometers.

It was found that the empirical formula

$$F = \frac{1000\sqrt{P_r}}{D^2} \quad (D > 600 \text{ km, winter day}) \tag{1}$$

fits the winter data satisfactorily for distances greater than 600 kilometers. This formula is shown dotted in Fig. 4. It may be mentioned that the daytime field intensity of WLW at 644 kilometers and on 700 kilocycles does not change much with season while the daytime field intensity of WCKY at 688 kilometers and on 1490 kilocycles is at least 100 times as strong during the winter day as for the summer day. It may be mentioned that the sky-wave formula (1) agrees with the Sommerfeld ground-wave formula at great distances for values of conductivity $\sigma_{emu} = 4f^2 \cdot 10^{-20}$ where f is the frequency in kilocycles; i.e., for

a conductivity of 10^{-14} electromagnetic units at 500 kilocycles or a conductivity of 10^{-13} electromagnetic units at 1600 kilocycles.

In Fig. 5 the measured night field intensities are shown for about three hundred stations as a function of the distance from the transmitter. Most of these measurements were made after 2 A.M. during the Federal Radio Commission's frequency monitoring schedules. Each point represents the maximum during a ten-minute time interval or the average of several such maxima but never more than six corresponding to an hour. There is one point for each separate night that

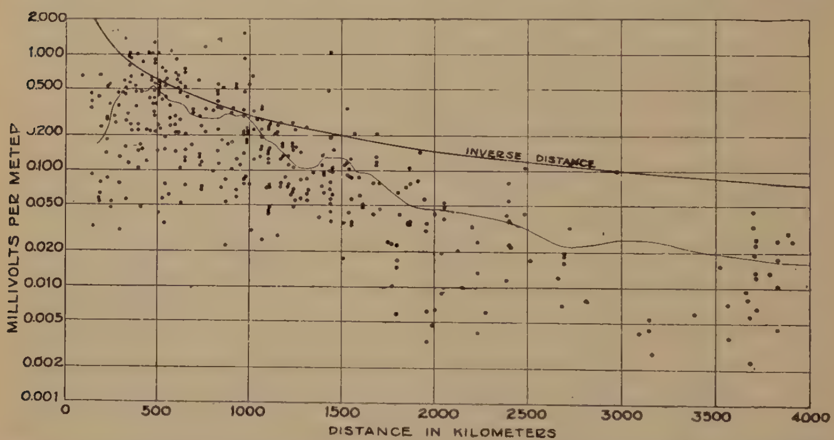


Fig. 5—Night field intensity measurements reduced to one kilowatt of radiated power and given as a function of distance for the broadcast stations in the United States; the irregular curve represents a running average of the data.

the station was observed. They were corrected for power so as to correspond to one kilowatt of radiated power. It will be observed that many of the maxima exceed the calculated inverse distance field as was evident also on the diurnal characteristics. Some of this excess is possibly due to too low an estimate of the radiated power but it is believed that part of it is real and may be explained on the assumption that several sky waves combine to give the observed maxima. The three points at a distance of 7775 kilometers correspond to two stations in Honolulu, Hawaii. The irregular curve shows the variation of night field intensity with distance as determined by a running average[3] of

[3] The data were ordered according to distance. The field intensities for the first ten distances were averaged to give an average field intensity for the distance corresponding to the average of the first ten distances. The next average value of field intensity was determined by averaging the fourth to the thirteenth measured value of field intensity, etc. Since one application of this moving average process did not result in a sufficiently smooth curve, the process was repeated once to give the curve plotted in Fig. 5.

the data for distances less than 4000 kilometers. The general shape of this curve with a rise between 400 and 1000 kilometers is characteristic of broadcast transmission at night. The data obtained during the months of December, 1933, January, February, and April, 1934, were analyzed by months and no significant difference was found. However, as summer approached, the measurements were less reliable on account of the high atmospheric noise level, consequently small differences could not be determined. Continuous twenty-four-hour records of the field intensity as received at Meadows, Md., of station WEAF in New York, are available for a period of over a year. Similarly records are

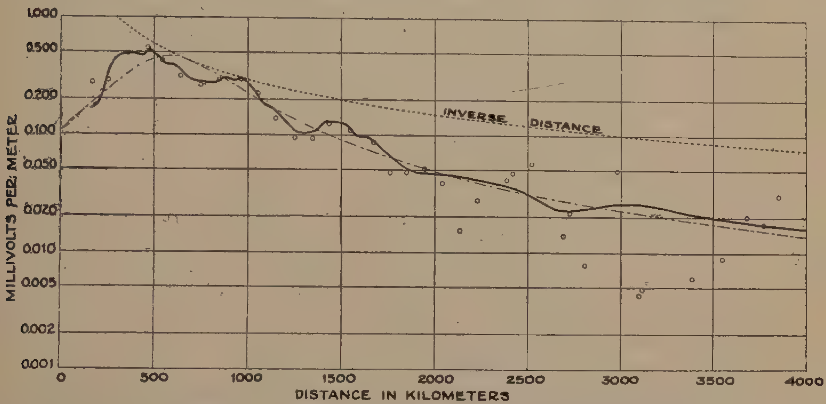


Fig. 6—Two averages of night field intensity measurements together with the graph of an empirical formula which was designed to represent the data.

available for WCKY in Covington, Ky., for over a year and for WLW for a period of greater than six months. These records are now being analyzed to determine samples of the seasonal variations at broadcast frequencies. Similarly, the data were analyzed for four different frequency ranges in the broadcast band and no significant changes with frequency were found. Thus it may be concluded that there is no large variation of night field intensity with season or frequency for the range from 550 to 1500 kilocycles.

It was found that the empirical formula

$$F = C\sqrt{P_r}\,\frac{1600 + D}{4{,}410{,}000 - 11{,}000D + 10D^2} \text{ (night sky wave)} \qquad (2)$$

fits the averaged data satisfactorily. This formula, together with the average data, is shown in Fig. 6. The maximum sky-wave field intensity predicted by (2) occurs at 580 kilometers.

The points plotted in Fig. 6 were obtained by an independent method of averaging and give credence to the major humps at 500,

1000, and 1500 kilometers which were exhibited by the running average of the data. Each point represents the average of the measured field intensities in a hundred-kilometer interval of distance plotted at the average distance in the interval. The 0–100, 3200–3300, and 3400–3500-kilometer intervals contained no measurements; the maximum number of measurements per 100-kilometer interval was 30 which were made in the 1100–1200-kilometer interval.

II. A Theoretical Explanation of the Data

Two simple theories of the attenuation of radio waves in the ionosphere were developed. The details of these theories are given in the appendix. The attenuation was calculated from these theories and an average polar diagram (i.e., $\sqrt{P'_r(\psi_1)}$, equation (23)) was then calcu-

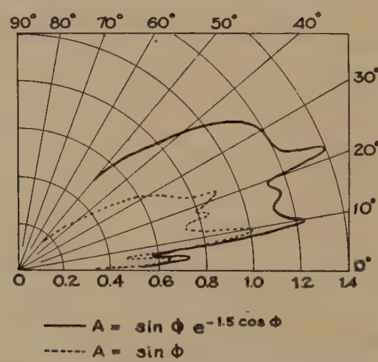


Fig. 7—Theoretical average polar radiation characteristics for broadcast stations in the United States.

lated from the averaged field intensity data given in Figs. 5 and 6. Fig. 7 shows two polar diagrams thus calculated. It should be emphasized that these represent the average polar characteristic of all the transmitting antennas multiplied by the polar characteristic of the receiving antenna. The receiving antenna was a vertical wire less than a quarter wavelength long. The dotted graph was determined by setting $K=O$ in (23) on the assumption that the waves were reflected without absorption at a layer one hundred kilometers high; i.e., on the assumption that the attenuation was due only to the fact that the wave traveled to a hundred-kilometer layer and back. This assumption corresponds for the purpose of our analysis to that used by P. P. Eckersley[4] and later by Stuart Ballantine[5] and gives an attenuation factor $A=\sin\phi_1$ where ϕ_1 denotes the angle of incidence at the layer.

[4] Proc. I.R.E., vol. 18, pp. 1160–1194; July, (1930).
[5] Proc. I.R.E., vol. 22, p. 619; May, (1934).

(See Appendix I.) The amount of radiation above 45 degrees (see dotted curve, Fig. 7) which is calculated on this assumption is hardly as much as might be expected from the average transmitting antenna in use by the stations measured. The solid curve in Fig. 7 was calculated from the average data on the assumption that the waves were refracted back to earth at a hundred-kilometer layer in which the ion density varied exponentially with the height. This assumption gives an attenuation factor independent of the frequency and equal to $\sin \phi_1 e^{-k \cos \phi_1}$. (See Appendix II.) The constant k was arbitrarily taken to be 1.5. It is evident that a larger amount of high angle radiation is obtained with an attenuation factor of this form and the relative amount may be changed further by an adjustment of the constant k. Since the solid curve in Fig. 7 has approximately the general shape which might be expected for $\sqrt{P_r(\psi_1)}$, we are led to believe that (23) properly relates the variables influencing sky-wave propagation in the broadcast band at night. The constant k, being theoretically proportional to the average collision frequency of the ions along the ray path, may very reasonably be expected to vary in a large measure from night to night since small changes in k might cause large changes in the pressure and thus in the collision frequency. This will explain the fact that field intensity observations on the same station at a fixed distance vary enormously from night to night. This probable variation in k may also be offered as a partial explanation of the large variation of the data in Fig. 5 above and below the averaged data; the remainder of this variation is probably due to a variation of $P_r(\psi_1)$ for the different stations measured.

For distances greater than 2000 kilometers the field intensities are received after two or more reflections since the curvature of the earth shields the receiver from the first reflection. At these distances also the refraction theory gives a more adequate explanation of the data, the simple reflection theory tending to predict too large values of received field intensity.

After the recent installation of the new 500-kilowatt transmitter for WLW, their engineers made many field intensity measurements. Fig. 8 shows the results of several sets of measurements after they were corrected for radiated power. Each point represents the maximum field intensity observed at a given location. The period of observation for each of these points varied from about five minutes to about five hours. A loop antenna was used for reception. These data are in good agreement with our own at the greater distances. The low values observed between 300 and 500 kilometers are probably caused by the low angle radiator (0.58λ vertical antenna) used by this station. Thus in the sin-

gle case where we have measurements at various distances on a station for which we know the radiated power we obtain good agreement quantitatively with our averaged data or empirical formula (2) and qualitatively with our theory.

In conclusion we may enumerate the deductions made from the analysis of the data. Received night field intensities are often greater than inverse distance, reaching their maximum values at about 600 kilometers. The average night field intensity for the average broadcast station may be predicted by means of (2) but the individual measurements on a station may be expected to vary by a factor of 10 above or below its average. There are no large variations of received field inten-

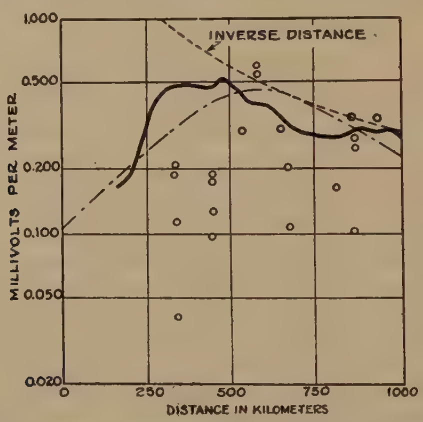


Fig. 8—Night field intensity measurements of WLW compared to averaged data for other broadcast stations in the United States.

sity with season or frequency although, in general, the received field intensity is slightly weaker in the summer. The variation of the received field intensity at night with distance appears to be determined primarily by the directional characteristics in the vertical plane of the transmitting and receiving antennas. The theoretical formula (23) seems to give an adequate explanation of the data, the variation of $P_r(\psi_n)$ giving an adequate explanation of the variation of received field intensity with distance and the variation of k giving a plausible explanation of the night-to-night and diurnal variations. Field intensities received during the daytime at distances greater than about 600 kilometers are sky waves and may be predicted during the winter time by (1).

Appendix I. Reflection

It is assumed in deriving this theory for the attenuation of sky waves at broadcast frequencies that the waves are reflected or refracted

back to earth from a layer in the ionosphere of virtual height 100 kilometers. The term virtual height is defined in papers by G. Breit and M. A. Tuve,[6] and T. R. Gilliland, G. W. Kenrick, and K. A. Norton.[7] The theory is applied by making the further assumption that the primary portion of the energy reflected back to earth reaches the receiver after only a single reflection at the layer for distances less than about 2000 kilometers and after only two reflections for distances greater than 2000 but less than 4000 kilometers.

Fig. 9 shows the geometry for one of the n hops between the transmitter and receiver. Let ϕ_n denote the angle of incidence of the waves

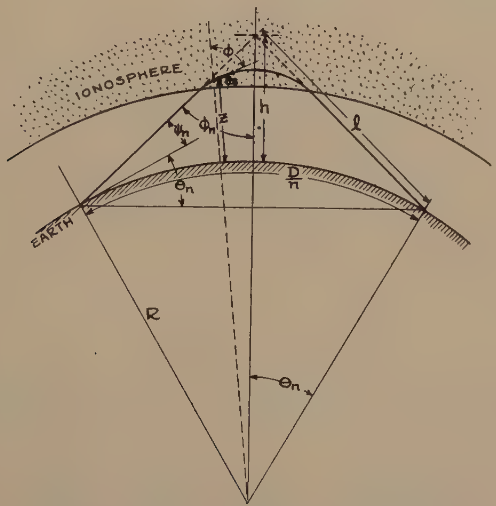


Fig. 9—Diagram indicating the geometry and the parameters concerned in theory of sky-wave propagation.

at the layer, $2n\theta_n$ the angular distance between the transmitter and the receiver, and ψ_n the angle above the horizon with which the waves leave the transmitter or arrive at the receiver. Here n denotes the number of reflections at the layer which the waves make before arriving at the receiver. Let $R=6371$ kilometers denote the radius of the earth, $h=100$ kilometers denote the virtual height of the layer, and l the distance from the transmitter to the layer. The geometry of Fig. 9 gives

$$\tan \phi_n = \frac{\sin \theta_n}{1 - \cos \theta_n + \dfrac{h}{R}} \tag{3}$$

[6] *Phys. Rev.*, vol. 21, p. 554, (1926).
[7] *Bur. Stan. Jour. Res.*, vol. 7, December, (1931), (RP390); Proc. I.R.E., vol. 20, pp. 286–310; February, (1932).

$$\psi_n = \frac{\pi}{2} - \phi_n - \theta_n. \tag{4}$$

If, as before, D denotes the distance in kilometers between transmitter and receiver we have

$$D = 222.1 n\theta_n \text{ degrees}. \tag{5}$$

The greatest distance for which waves may be reflected back to earth after a single reflection is determined by setting $\psi_1 = 0$ in (4) and substituting the resulting value of ϕ_1 in (3), whence,

$$\theta_{max} = \cos^{-1} \frac{R}{R+h} = 10°5'. \tag{6}$$

In Fig. 10 ϕ_n and ψ_n are given graphically for one and for two reflections at the layer as a function of D. In general for n reflections we have the relations

$$\phi_n(D) = \phi_1\left(\frac{D}{n}\right) \tag{7}$$

$$\psi_n(D) = \psi_1\left(\frac{D}{n}\right). \tag{8}$$

If we assume that the waves are reflected at the layer without absorption, then they will be received with an intensity

$$F = \frac{C}{D}\sqrt{P_r(\psi_n)}A_n = \frac{C}{2nl}\sqrt{P_r(\psi_n)} \tag{9}$$

so that,

$$A_n = \frac{D}{2nl} = \frac{\theta_n}{\sin\theta_n}\sin\phi_n \cong \sin\phi_n \tag{10}$$

where the attenuation factor A_n simply gives the reduction in field intensity due to traveling the inverse path distance as compared to the inverse great-circle distance.

It is possible to neglect the ratio $\theta_n/\sin\theta_n$ since θ_{max} for a hundred-kilometer layer is only $10°5'$. Equation (10) then represents the attenuation factor for a single wave at a reflecting layer and is applicable for any finite distance between transmitter and receiver and for any number of reflections. It is evident that the succeeding reflections after the first will be weaker in the ratio $\sin\phi_n/\sin\phi_1$ and will be neglected in this discussion.

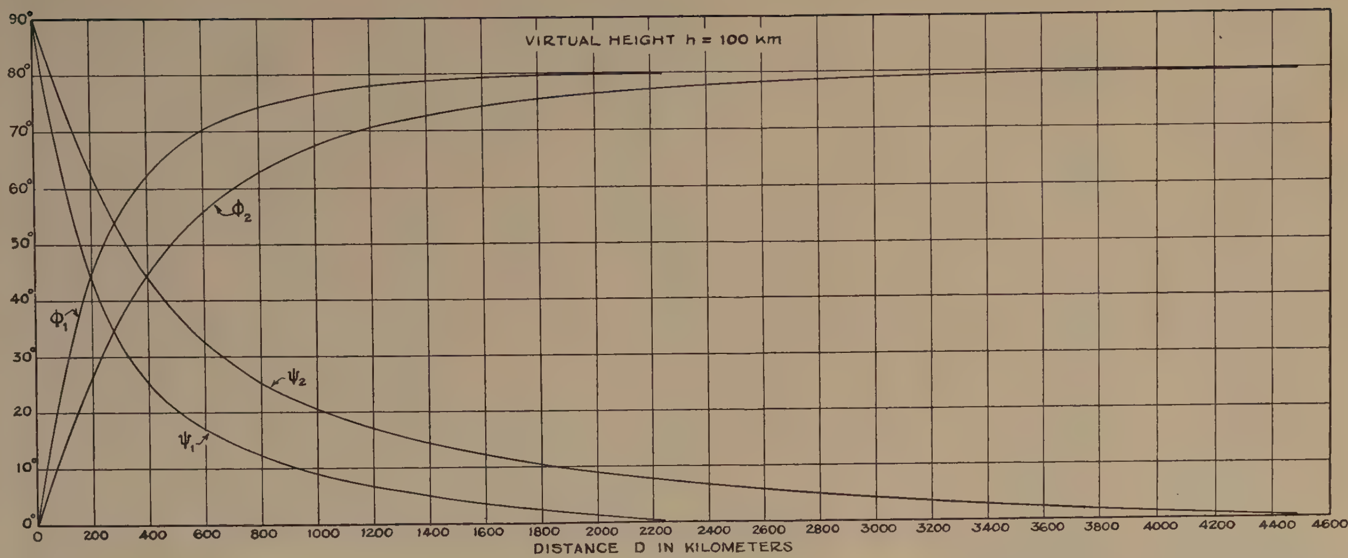


Fig. 10—Diagram indicating the variation with distance of two important parameters in the theory of sky-wave propagation ϕ, the angle of incidence at the layer, and ψ, the angle above the horizon at which the radiation from the transmitter takes place.

Appendix II. Refraction

Formula (10) was determined on the assumption that the waves were reflected without absorption at the layer. If we set Γ_n equal to the total attenuation along the ray path and γ equal to the attenuation per unit distance, then,

$$\Gamma_n = \int \gamma dS \tag{11}$$

where S denotes the distance along the ray path and the integration extends along the entire path. Now from the classical electromagnetic theory[8] we have the relations

$$\gamma \sim \frac{Nv}{\mu\omega^2} \tag{12}$$

$$\mu^2 = 1 - \frac{CN}{\omega^2} \tag{13}$$

$$\cos\psi_n = \sin\phi\mu\left(1 + \frac{Z}{R}\right) \tag{14}$$

where N denotes the ionic density, v the collision frequency of the ions, μ the index of refraction, ω the angular frequency, Z the distance above the surface of the earth at any point along the ray path, and C a constant. Results of ionization measurements during the solar eclipse of 1932 as reported by S. S. Kirby, L. V. Berkner, T. R. Gilliland, and K. A. Norton[9] indicate that the ionization during the daytime in the E layer consists mostly of heavy ions so that the modification of (13) due to the magnetic field of the earth may be neglected. At night this is probably no longer strictly true so that this constitutes one of the approximations of the theory. Equation (14) denotes the invariant for the ray path around a curved earth with ϕ denoting the angle which the ray makes with the vertical. From (3) and (4) we obtain

$$\cos\psi_n = \cos\left(\frac{\pi}{2} - \phi_n - \theta_n\right) = \sin\phi_n\left(1 + \frac{h}{R}\right). \tag{15}$$

Combining (14) and (15) we obtain

$$\sin\phi_n = \mu\sin\phi\,\frac{R + Z}{R + h} \cong \mu\sin\phi. \tag{16}$$

[8] *See*, e.g., P. O. Pedersen "The Propagation of Radio Waves," Copenhagen, 1927.

[9] *Bur. Stan. Jour. Res.*, vol. 11, p. 829; December, (1933), (RP629); Proc. I.R.E., vol. 22, pp. 247–265; February, (1934).

Since there will be no absorption except in the layer, Z will not be much less than h throughout the range of integration so that we may safely make the approximation of setting the ratio $(R+Z)/(R+h)=1$. The ray reaches the highest point and the greatest ion density, say N_o, when $\sin\phi=1$ so that we obtain from (13) and (16)

$$N_0 = \frac{\omega^2}{C}\cos^2\phi_n. \tag{17}$$

At any point along the ray path we have

$$dS = \frac{dZ}{\cos\phi} = \frac{\mu dZ}{\sqrt{\cos^2\phi_n - \frac{CN}{\omega^2}}} = \frac{\mu dZ}{\cos\phi_n\sqrt{1-\frac{N}{N_0}}}. \tag{18}$$

Substituting (12) and (18) into (10) we obtain:

$$\Gamma_n \sim \frac{n}{\omega^2\cos\phi_n}\int\frac{NvdZ}{\sqrt{1-\frac{N}{N_0}}} = \frac{n}{\omega^2\cos\phi_n}\int_0^{N_0}\frac{NvdN}{\frac{dN}{dZ}\sqrt{1-\frac{N}{N_0}}} \tag{19}$$

the first integration extending from the height for which $N=O$ to the height for which $N=N_0$. We will use the theorem of the mean to replace v by its mean value $\bar{v}$ and assume that N varies exponentially with the height so that

$$\Gamma_n \sim \frac{n\bar{v}}{\omega^2\cos\phi_n}\int_0^{N_0}\frac{dN}{\sqrt{1-\frac{N}{N_0}}}. \tag{20}$$

Using the new variable $X=N/N_0$ we obtain

$$\Gamma_n \sim \frac{n\bar{v}N_0}{\omega^2\cos\phi_n}\int_0^1\frac{dX}{\sqrt{1-X}} \sim n\bar{v}\cos\phi_n. \tag{21}$$

Thus we obtain the result that the attenuation along the ray path[10] is proportional to the product of the number of reflections, the average collision frequency of the ions and the cosine of the angle of incidence at the layer. It is important to notice that the attenuation for the assumed exponential distribution of ion density is independent of the frequency. Since our results indicated no variation of received field intensity with frequency and since an exponential distribution of

[10] This same problem was solved in terms of somewhat less convenient variables in a paper by Shogo Namba, PROC. I.R.E., vol 21, pp. 238–263; February, (1933).

ion density is very probable because of the exponential absorption in the ionosphere of ultra-violet radiation from the sun, (21) very likely has the correct form. Combining the attenuation effects due to (10) and (21) we obtain the general expression for the attenuation factor for a ray after n reflections at the layer

$$A_n = \sin \phi_n e^{-nk \cos \phi_n} \tag{22}$$

and the field intensity of this wave will be

$$F_n = \frac{C}{D}\sqrt{P_r(\psi_n)} \sin \phi_n e^{-nk \cos \phi_n} \tag{23}$$

where the radiated power is for the angle $\psi_n(D)$ and k is a constant independent of D and f but proportional to the average collision frequency of the ions along the ray path.

Proceedings of the Institute of Radio Engineers
Volume 23, Number 10 *October, 1935*

NEGATIVE RESISTANCE AND DEVICES FOR OBTAINING IT*

By
E. W. Herold
(RCA Radiotron Company, Inc., Harrison, New Jersey)

Summary—*By the use of a concept due to Crisson, negative resistance is shown to be a phenomenon controlled by either current or voltage but not by both together. Thus, two main classes are found differing in the shape of the volt-ampere characteristic, in the conditions for stability, in the effect of an added positive resistance, and in the effect of an internal time lag. Reliability, good power conversion efficiency, and a low ratio of the unavoidable self-reactance to the negative resistance are mentioned as being desirable characteristics of any negative resistance. A figure of merit for the voltage controlled type is shown to be* $1/\omega C_v R_v$, *where* C_v *is the unavoidable effective shunt capacitance. In the current controlled case* $R_c/\omega L_c$ *is shown to be a merit factor,* L_c *being the effective series inductance. The addition of external positive resistance lowers the figure of merit for the combination.*

A number of well-known devices are discussed after classifying them in three groups according to principle of operation. The simple group includes such devices as the dynatron and arc which produce negative resistance between two elements. The direct coupled group comprises primarily vacuum tubes having negative transconductance in which negative resistance is produced by a direct connection between the controlling and controlled electrodes. The general properties of the negative transconductance tube as a negative resistance are detailed, including the effect of interelectrode capacitances. The third group, the reverse phase coupled group, is included but not treated in detail. It comprises vacuum tube arrangements requiring a phase reversing means for coupling the controlling to the controlled electrodes.

Applications of negative resistance to the production of sinusoidal and relaxation oscillations, to circuits having special properties, and to measurement work, are mentioned with frequent reference to the bibliography. In addition, the features of a tube having negative transconductance as an amplifier are pointed out.

In conclusion a section is devoted to the characteristics of the type 57 tube which has negative transconductance between the third and the second grid. As a voltage controlled negative resistance, this tube will produce a negative resistance of 3500 ohms with a total cathode current of only seven milliamperes. This performance is believed to be better than that of most dynatrons.

Introduction

NEGATIVE resistance has held considerable interest in radio communication since Hull first disclosed the dynatron. In spite of the fact that many other arrangements permit the production of negative resistance, the words "negative resistance" immediately suggest the dynatron. A study of the field, however, indicates that a

* Decimal classification: R139. Original manuscript received by the Institute, September 21, 1934. Presented in part before Washington Section, April 9, 1934.

large number of usable devices produce this effect, and some of them have qualities in some respects superior to those of the dynatron. It is the purpose of the following discussion both to indicate the fundamental physical principles of negative resistance and to emphasize the usefulness of negative resistance in communication work. In particular, it is hoped to emphasize the advantages of some of the less well-known methods of obtaining negative resistance.

Although the words "negative resistance" indicate the opposite of positive resistance, devices which possess it are necessarily limited in the energy which they can handle, so that they exhibit negative resistance symptoms only over a limited range, a property not ordinarily associated with positive resistance. A positive resistance dissipates energy proportional to the square of the impressed voltage or current, whereas a negative resistance *generates* energy proportional to the square of the impressed voltage or current. In order to generate energy proportional to an impressed electrical disturbance, a negative resistance must have a source of energy and some means of controlling it. Crisson[1] presented a useful picture of how the generated energy of a two-terminal device may be made proportional to the square of the impressed voltage or current. His picture was in terms of an ideal amplifier, such as shown in Fig. 1(a), where an impressed electromotive force at the input terminals generates a proportional voltage or current at the output terminals. If the input terminals are connected to the output terminals, as in Fig. 1(b), it is now evident that the output of the amplifier becomes proportional to the voltage across the output terminals. In order to obtain negative resistance, the connections must be made in such a way that instability will result unless the voltage is limited by the connection of an external resistance of a low enough value across the terminals. The operation of such an arrangement when a small impressed electromotive force (represented by the generator in the figure) is considered, is such that the voltage across the terminals causes the amplifier to send a proportional current through the external circuit in such a direction that its IR drop through the external resistance adds to the original impressed voltage. As the external resistance is increased, the voltage added to the original disturbance is increased, until the external resistance becomes so high that the IR drop in this resistance becomes greater than the controlling voltage needed to produce the current I, and the regenerative effect causes instability. This condition occurs when the external resistance is greater than the negative resistance. The negative resistance produced by the amplifier is indicated by the ratio of the voltage incre-

[1] Numbers refer to bibliography.

ment to the current increment across the terminals. The greater this ratio, the higher the negative resistance, and thus, the greater the voltage that is required to produce a given current. Therefore, we may say qualitatively that in this case a low value of negative resistance is desirable. Another way of producing negative resistance is to connect the input terminals of the amplifier across a resistance in series with the amplifier output, such as is shown in Fig. 1(c). In this case the amplifier input, and hence its output, is proportional to the *current* through the circuit rather than to the voltage across it and the behavior is quite different. Instability will now result when a sufficiently

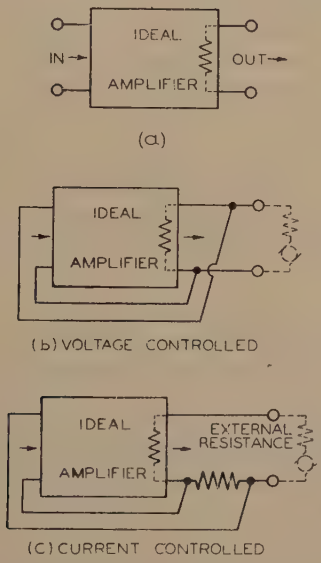


Fig. 1—Ideal amplifier used to illustrated conception of negative resistance.

low external resistance is present. The operation with an impressed disturbance is now such that the current through the circuit causes the amplifier output voltage to *oppose* the *IR* drop in the external resistance. The original current is increased by this effect. As the external resistance is decreased, the current is increased until the external resistance is so low that the regenerative effect causes instability. The unstable condition occurs when the external resistance is less than the negative resistance. The negative resistance produced is again indicated by the ratio of voltage-to-current increment but in this case, since current control is used, a high ratio or a high negative resistance is desirable so that the maximum voltage is generated for a given current.

The two methods of producing negative resistance which have been described are illustrative of a fundamental principle underlying all negative resistance devices. A physically realizable negative resistance is not the reverse of a positive resistance (in which the current-voltage relationship is an interdependent one) but utilizes an internal phenomenon which depends only on the voltage or only on the current. The words "negative resistance" are, therefore, somewhat misleading but their widespread use discourages an attempt to introduce a new name. Throughout the following discussion, devices will be repeatedly referred to as voltage controlled negative resistances and as current controlled negative resistances, since it is into either of these two classes that all arrangements must be placed. To summarize, a negative resistance, as the term is here used, comprises a two-terminal device with an internal source of energy which is controlled either by the current through or by the voltage across these terminals but not by both.

Historically, Barkhausen[2] was one of the earliest investigators to indicate the fundamental difference in behavior of the two types. The first clear and conclusive demonstration of the universal validity of the current or voltage control principle seems to have been made by Steimel.[3]

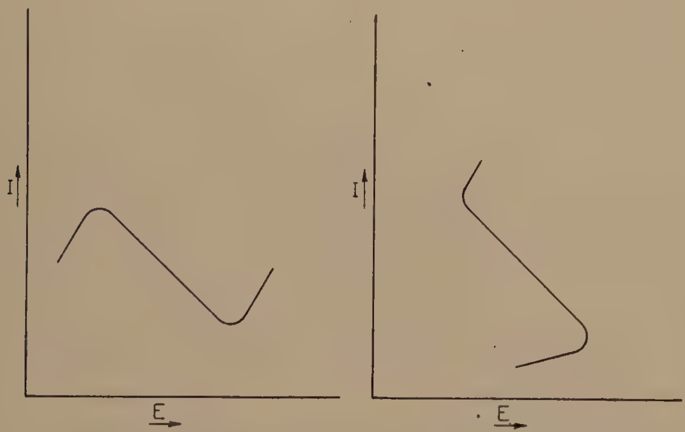
Characteristic Differences between the Two Classes of Negative Resistance

Since it is found that the voltage controlled and current controlled negative resistances behave differently when inserted in a circuit, it may be useful to list some of the major differences between the two from the circuit point of view.

1. Curve Shape

The volt-ampere characteristic of a negative resistance would be expected to have a negative slope over an appreciable range. The region of negative slope is fundamentally limited by the limited source of energy available; a second limitation is imposed by whatever controls the source of energy and enables the negative slope to occur. In the characteristic of a voltage controlled device (a typical one is shown in Fig. 2(a)), the limitations cause the negative resistance at the ends of the straight part to increase passing through infinity and becoming positive. This is clearly the result of the voltage control, since towards the ends of the negatively sloping part, a small increment in voltage fails to produce the change in current which it would produce were the increment applied to the center of the characteristic. In the typical volt-ampere characteristic of a current controlled negative resistance,

as shown in Fig. 2(b), the limitations cause a smaller voltage change for a given current increment at the ends of the straight part and the negative resistance at the ends passes through zero and becomes positive. It should be noticed from the figures that phenomena observed with the resistance in the one case will be observed with the conductance in the other, since an interchange of the current with the voltage axes of either of the two, causes the characteristics to be similarly shaped. This fact is most important because it enables the immediate converting of all phenomena established for the one type of negative resistance, to analogous phenomena for the other.



(a) VOLTAGE CONTROLLED (b) CURRENT CONTROLLED

Fig. 2—Falling characteristics exhibited by negative resistance devices.

2. Conditions for Stability

An unstable point of operation may be said to be one in which an infinitesimal fluctuation of current or voltage is sufficient to cause the point of operation to shift automatically to another point, usually a stable one. The voltage controlled negative resistance operates at a stable point when the external positive resistance is less than the negative resistance at the point. This was established in connection with Fig. 1(b). Use of the resistance-conductance analogy mentioned under *Curve Shape* indicates that the current controlled negative resistance operates at a stable point when the external positive conductance is less than the negative conductance at the point; i.e., the external positive resistance must be greater than the negative resistance at the point.* This has been already established in connection with Fig. 1(c).

* In order to measure the volt-ampere characteristic of a current controlled negative resistance such as the one shown in Fig. 2(b), it is necessary to insert a high stabilizing resistance in series with the supply battery.

3. Conditions for Excitation of Oscillation in a Tuned Circuit

A voltage controlled negative resistance is suitable for exciting oscillations in a circuit which has a high resistance at resonance; e.g., a parallel tuned circuit of L,C, and r.[3,4] The condition for oscillation is that the negative resistance be less than L/Cr. A current controlled negative resistance will similarly excite a circuit which has a high conductance (low resistance) at resonance; e.g., a series tuned circuit.[3,] The condition for oscillation is that the negative resistance be greater than the positive resistance of the circuit. Use of the wrong combination of circuit and negative resistance results in parasitic or relaxation oscillations, if any occur at all.[3]

4. Conditions for Excitation of Degenerated Sinusoidal or Relaxation Oscillations[5,6]

If the capacitance of a parallel tuned circuit is made vanishingly small and the resistance of the inductance coil is sufficiently low, the connection to the usual voltage controlled negative resistance with limited negative slope may result in oscillations whose period depends on a relaxation time. Similarly, if the inductance of a series tuned circuit connected to a current controlled arrangement is made vanishingly small, relaxation oscillations may occur. In both cases the periodicity of the oscillations is usually a function of the time constant of the complete circuit.

5. Effect of Added Positive Resistance

If a voltage controlled negative resistance is not sufficiently small to excite a particular tuned circuit, the addition of series positive resistance will lower the effective negative resistance.[7] Analogous reasoning indicates that positive resistance shunted across the current controlled device will raise the negative resistance. The possibilities of these methods are seriously limited by the unavoidable self-reactance inherent in a negative resistance devices. A more detailed treatment will be given in a later section.

6. Effect of Time Lag between Cause and Effect

In many negative resistance arrangements, the control mechanism introduces an unavoidable time lag between the application of the control factor and the resulting output. The effect causes the dynamic volt-ampere characteristic to become a loop departing noticeably from the static characteristic. For small time lags, in the voltage controlled case, the sense of rotation of the loop is clockwise and the effect of the circuit may be considered as the result of a negative resistance without lag shunted by a capacitance. An example is given in Fig. 3(a).

The loop in the current controlled case has a sense of rotation which is counterclockwise so that the circuit effect is that of a negative resistance without time lag in series with an inductance. An illustration of the effect in the current controlled case is shown in Fig. 3(b). Since the values of the effective capitance introduced in the one case and the effective inductance in the other are functions of the frequency, the effects cannot be eliminated and the behavior must be considered as a function of frequency.

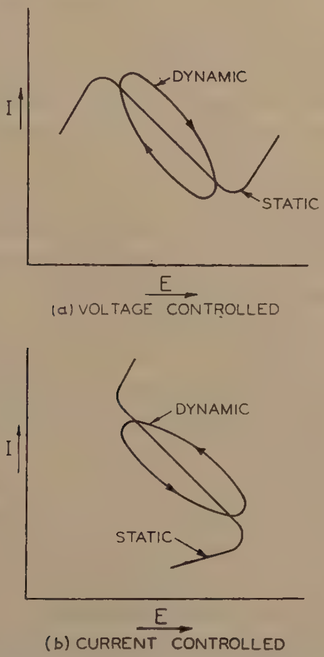


Fig. 3—Dynamic loops occurring with short time lag between cause and effect.

Desirable Characteristics of a Negative Resistance

The desirable characteristics of a negative resistance, just as in the case of other amplifying arrangements, are dependent to a great extent upon the use to which the device is to be put. For most general use, a negative resistance should be reliable, convenient, and simple to use, and should have reasonably good efficiency in transforming supplied power into useful power. The ability to produce either voltage controlled or current controlled negative resistance by a difference in connections is possessed by some arrangements and is advantageous for general usefulness. Qualitatively, it may be stated that with a voltage

controlled device, a low negative resistance is desirable since this indicates a large change in current for a given control voltage. In the current controlled case, a large change in voltage for a given control current, or a high negative resistance, would appear to be desirable.

An attempt to go further into the characteristics which a negative resistance should have leads to the consideration of the inherent self-reactance of the devices. For most general usefulness, it is evident, of course, that the unavoidable self-capacitance or self-inductance of a negative resistance should be as small as possible. The capacitance or inductance which must be considered should include the effective capacitance or inductance introduced by the time lag, if any. It is difficult to evaluate quantitatively the effect of the reactance without consideration of the particular application of the negative resistance to be made. Certain points, however, stand out in their generality of application.

The reactive part of the impedance of a voltage controlled negative resistance almost always takes the form of shunt capacitance only. If a voltage controlled negative resistance, R_v, and a shunt capacitance, C_v, are considered, the ratio of the reactance to the resistance at an angular frequency, ω, is $\omega C_v R_v$. In most applications, it is found that $\omega C_v R_v$ is a factor of poorness. Its reciprocal $1/\omega C_v R_v$, therefore, is a figure of merit. In these applications two identical voltage controlled devices in parallel are no better than one except as regards the power available.

In the current controlled case, the reactance present may be considered in the form of inductance, L_c, in series with the negative resistance, R_c, with an additional complicating capacitance shunting the whole in some cases. If the latter is forgotten for the moment, analogous reasoning to that used in the voltage controlled case leads to a poorness factor, $\omega L_c/R_c$, the ratio of reactance to resistance. A figure of merit, therefore, would be $R_c/\omega L_c$. The capitance shunting the whole, as already indicated, may result in relaxation oscillations if high external resistance is present, and should, therefore, be very low. Preferably $\omega C R_c$ should be less than unity at the highest frequency at which negative resistance occurs. Particular cases may be quantitatively analyzed as regards this capacitance.*

Analysis shows that the addition of series positive resistance to a voltage controlled negative resistance, although it lowers the negative resistance,[7] lowers the figure of merit, $1/\omega C_v R_v$. It is easily shown that reducing the negative resistance, R_v, by this means to a value R_v/m increases the effective capacitance of the combination to $m^2 C_v$. Analo-

* For example, see the later section on direct coupled negative resistances.

gously in the current controlled case, shunt positive resistance raises the negative resistance but lowers the figure of merit, $R_c/\omega L_c$. In this case, raising the negative resistance to mR_c raises the effective inductance of the combination to m^2L_c.

Classification of Negative Resistance Devices

All methods of producing negative resistance may be placed into one of three classes according to the principle of operation. The first class includes the simple negative resistances or those in which the element exhibiting negative resistance also controls the internal phenomenon making the negative slope possible. The dynatron[8] is a well-known example. The voltage on the dynode controls its secondary emission and hence its current. The second class may be called the direct coupled negative resistance group since it comprises those devices in which the element controlling the internal phenomenon must be directly connected to the circuit or to the controlled element before negative resistance is possible. An example is the negatron of Scott-Taggart[9] in which a rising control element potential causes a decreasing current to an anode. If the two are directly connected, negative resistance is obtained. The third class will here be called the reverse phase coupled negative resistance group since in it are found those devices in which the controlling element is again separate from the controlled element but must be coupled to it through a phase reversing means before negative resistance is obtained. The conventional vacuum tube with a ticker coil feed-back is an example.

Although the simple negative resistances are usually either current controlled or voltage controlled but not both, the devices in both of the coupled negative resistance groups may be used in either a current controlled or a voltage controlled arrangment depending on the way in which the control element is connected. This will be clarified by the detailed treatment of a subdivision of the direct coupled group to follow.

Simple Negative Resistance Devices

Among the many devices which are properly included under this heading, a few which are of some interest because of their value in communications will be listed.

1. *Arc or Glow Discharge,*[10] *Current Controlled**

A rising current in an arc or glow discharge causes such an increase in ionization that a smaller potential is required to maintain the discharge. Devices of this kind are not reliable because of changes in the

* Also voltage controlled under special conditions, Cf. Penning.[11]

electrodes and in the gas pressure. Because of the slow ion mobility, sufficient time lag is present to prevent successful high-frequency operation. The convenience and simplicity of the cold-cathode glow discharge has encouraged its use in low-frequency relaxation-oscillation circuits.

2. *Dynatron,*[8] *Voltage Controlled*

A rising voltage on an electrode in a high vacuum electron tube causes such an increase in secondary emission that the current to the electrode falls. A collector electrode for the secondary electrons placed at a high fixed potential is necessary. The dynatron suffers as a class from great variability of secondary emission sensitivity encountered with present-day exhaust technic.[12] Although low values of negative resistance are obtained only with high currents and voltages, the convenience of the dynatron has been an advantage in a wide variety of applications.[13]

3. *Field Decay Type,*[14,15] *Current Controlled*

This is a less well-known method of utilizing secondary emission to obtain negative resistance, in which a resistor connected between grid and plate of a triode allows the grid voltage to rise so high with the onset of secondary emission that the total voltage from plate to cathode necessary to maintain a current decreases as the current rises. It is believed that this device is inferior to the dynatron in reliability but it is of interest because it is current controlled.

4. *Space-Charge-Grid Type,* [16,17,18] *Voltage Controlled*

A rising potential on a space-charge grid causes an increased anode current so that if the cathode current is sufficiently limited, the space-charge-grid current drops. The limited emission necessary has been obtained in the past primarily by reduction of filament temperature,[17,18] leading to unreliable and unsatisfactory operation. The method has been used experimentally.

5. *Positive Ion Type, Voltage Controlled*

In a triode containing gas at low pressure, an increasing grid voltage increases the plate current and hence the positive ion current to the negative grid, thereby achieving negative resistance in the grid circuit.[19] This method is unreliable, has considerable time lag, and produces only very high values of negative resistance, but has been used to some advantage in direct-current amplifiers and in electrometer applications.

6. *Split-Anode Magnetron,*[20] *Voltage Controlled*

If in a cylindrical magnetron the anode is divided diametrically into two parts, and a suitable magnetic field is applied, a difference of potential between the two anode segments causes a decreased current to the more positive element and an increased current to the less positive. The tube, therefore, exhibits negative resistance. Although the method is reliable, it is somewhat inconvenient to use because of the magnetic field required. It has found chief application in high-frequency work where the very low internal capacitance is advantageous.

7. *Ionic Transit-Time Devices*

A time lag by electrons or positive ions comparable to the period of the impressed frequency is sufficient to cause an ionic device which exhibits positive resistance at low frequencies to appear as a negative resistance in certain high-frequency bands.[21,22,23] Too little is known about this phenomenon to draw definite conclusions as to its utility as compared with other means.

Direct-Coupled Negative-Resistance Devices

The most interesting devices in this group are those vacuum tube arrangements in which a rising potential on one element causes a decreasing current to another element. Since the observed operation of all tubes of this kind is similar, although perhaps the internal causes are different, it is well to examine the general aspects of the production of negative resistance by this means before listing some of the suitable devices. To fall in this group, a vacuum tube should have negative transconductance between two of its elements. The essential elements are a control element and an output anode electrode, any other electrodes present, although necessary to operation, playing little or no part in the external circuit operation. The tube may then be considered as a triode with negative transconductance; the amplification factor may be positive or negative depending on whether the anode resistance is negative or positive. In Fig. 4(a) is shown a modified triode symbol suitable for use with a negative transconductance tube. A typical transfer characteristic of such a tube is shown in Fig. 4(b). The equivalent circuit is, of course, the same as that of the usual triode and is shown in Fig. 4(c), where s_m denotes the normal control element to anode transconductance and s_n denotes the inverse, or anode to control element, transconductance. Although in the following analysis first order effects only are considered, if s_m, s_n, r_g, and r_p are considered not as constants but as functions of the amplitude, the conclusions are equally valid. Ordinarily, in the most useful cases, the control electrode

is biased negatively and, thus, draws no current. The inverse transconductance, s_n, is then zero. For the sake of generality, however, it has been included in the analysis.

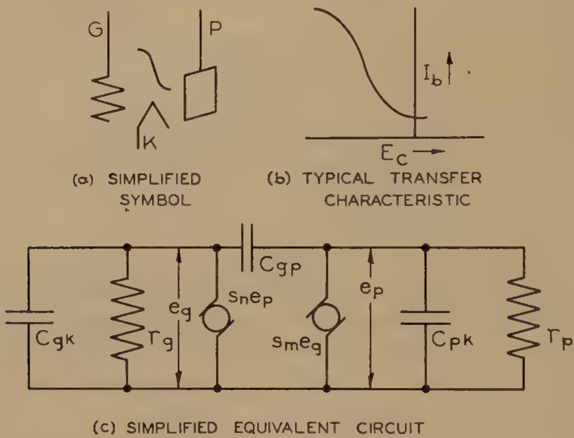


Fig. 4—Negative transconductance tube.

The connections which may be made to a negative transconductance tube to obtain voltage controlled negative resistance are shown in Figs. 5(a) and 5(b). In the former circuit, a suitable biasing battery

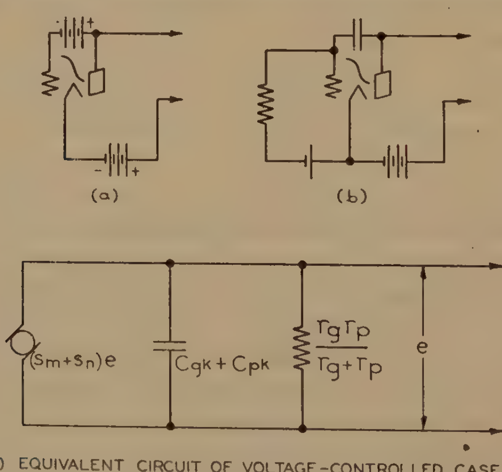


Fig. 5—Production of voltage controlled negative resistance.

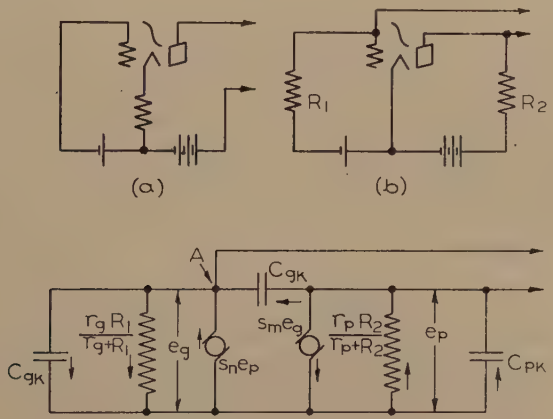
is connected directly between anode and control element. A rise in anode potential then causes an equal rise in control element potential, so that, if the tube parameters are of suitable value the total current

falls. An equivalent and more useful circuit for alternating-current phenomena is shown in the second circuit in which the control element and anode are connected through a large condenser and the control element bias is applied through a high resistance. The equivalent circuit of the first arrangement and of the second for a very large capacitance and very high control element external resistance is shown in Fig. 5(c). Inspection indicates that the arrangement is equivalent to a resistance

$$R_v = \frac{1}{s_m + s_n + \frac{1}{r_g} + \frac{1}{r_p}}$$

shunted by a capacitance

$$C_v = C_{gk} + C_{pk}.$$



(C) EQUIVALENT CIRCUIT OF CURRENT-CONTROLLED CASE

Fig. 6—Production of current controlled negative resistance.

If r_g and r_p are positive, $s_m + s_n$ must be negative and greater in magnitude than $1/r_g + 1/r_p$ for the resistance to be negative. It is, therefore, desirable that r_g, r_p, and s_m be high. It will be noted that the circuit, in common with other voltage controlled negative resistances, may be considered primarily as a capacitance and a resistance in parallel.

Two circuits suitable for the production of a current controlled negative resistance are shown in Figs. 6(a) and 6(b). In these circuits the control element potential is negatively proportional to the current in the circuit. A rise in current causes a decrease in control element potential so that, if the tube parameters are again suitable, the anode

current rises and the anode voltage drops. Thus, the rising current is accompanied by a falling voltage, and a current controlled negative resistance results. An analysis of the equivalent circuit of the tube and connections shown in Fig. 6(c) indicates that the circuit comprises a resistance in series with an effective inductance (a result of the effect of C_{gk} and C_{pk} on the operation) with the whole shunted by a capacitance, C_{gp}. The actual impedance at the terminals, neglecting C_{gp}, is

$$Z_c = \frac{Z_g + Z_p + Z_g Z_p (s_m + s_n)}{1 - s_m s_n Z_g Z_p}$$

where,

$$\frac{1}{Z_g} = \frac{1}{r_g} + \frac{1}{R_1} + j\omega C_{gk}$$

$$\frac{1}{Z_p} = \frac{1}{r_p} + \frac{1}{R_2} + j\omega C_{pk}.$$

The significance of these expressions is perhaps best understood by the simplification occurring when the frequency is low enough so that the phase angle of Z_g and Z_p is very small, $r_g \gg R_1$, $r_p \gg R_2$, and the inverse transconductance is neglected. In this case the arrangement appears as a resistance

$$R_c = R_1 R_2 \left(\frac{1}{R_1} + \frac{1}{R_2} + s_m \right)$$

in series with an inductance

$$L_c = - \left[C_{gk} R_1^2 (s_m R_2 + 1) + C_{pk} R_2^2 (s_m R_1 + 1) \right],$$

the whole shunted by the capacitance C_{gp}. Again, it may be seen that a high value for s_m is desirable.

As already mentioned, the presence of C_{gp} offers the possibility of oscillation when the external circuit has a very high resistance. The condition for oscillation is easily derived by application of Kirchoff's laws to the circuit of Fig. 6. It is convenient to consider R_1 as the combined resistance of the internal and external control-element resistances in parallel and to consider R_2 as including the anode resistance. If this is so, the total current entering the lower wire of Fig. 6 (the direction of the currents is denoted by the arrows) is

$$e_g \left(\frac{1}{R_1} + j\omega C_{gk} + s_m \right) - e_p \left(\frac{1}{R_2} + j\omega C_{pk} + s_n \right) = 0.$$

The total current leaving the point A in Fig. 6 is

$$e_g\left(\frac{1}{R_1} + j\omega C_{gk}\right) - s_n e_p + j\omega C_{gp}(e_g + e_p) = 0$$

which may be written

$$e_g\left(\frac{1}{R_1} + j\omega C_{gk} + j\omega C_{gp}\right) + e_p(j\omega C_{gp} - s_n) = 0.$$

The only solutions for these two equations are $e_g = e_p = 0$, or

$$\begin{vmatrix} \left(\frac{1}{R_1} + j\omega C_{gk} + s_m\right) & -\left(\frac{1}{R_2} + j\omega C_{pk} + s_n\right) \\ \left(\frac{1}{R_1} + j\omega C_{gk} + j\omega C_{gp}\right) & (j\omega C_{gp} - s_n) \end{vmatrix} = 0.$$

The second, of course, gives the self-oscillatory condition. Equating the imaginary and the real parts of the determinant to zero gives

$$C_{gp} = -\frac{C_{gk}R_1 + C_{pk}R_2}{R_1 + R_2 + (s_m + s_n)R_1R_2} = -\frac{C_{gk}R_1 + C_{pk}R_2}{R_c}$$

$$\omega^2 = \frac{1}{R_1R_{2c}(C_{gk}C_{gp} + C_{pk}C_{gp} + C_{gk}C_{pk})} - \frac{s_m s_n}{(C_{gk}C_{gp} + C_{pk}C_{gp} + C_{gk}C_{pk})}.$$

For oscillation not to exist with an open circuit at the terminals of the current-controlled negative-resistance circuit, C_{gp} must be less than the value given above. This indicates clearly that for a tube to be capable of producing a high negative resistance of the current controlled type, C_{gp} must be much smaller than the other capacitances of the tube.

The results of all the considerations indicate that a good negative transconductance tube should have high transconductance, high anode resistance, high control element resistance, and low capacitances, in particular the control element to anode capacitance, if current controlled negative resistance is to be used.

1. *The Negatron*[9]

In this tube the current from a filamentary cathode is diverted from one to another of the two anodes on opposite sides of the filament by the potential applied to a grid interposed between the filament and one of the anodes. A rising grid potential reduces the current to the opposite anode, provided the filament is operated at a low enough temperature to be emission saturated. The latter requirement makes

the negatron unreliable and inconvenient. In addition, only low transconductance is achieved.

2. *Retarding Field Tubes*

In a tube with a cathode surrounded by a positive grid which in turn is followed by a low potential element or group of elements, a rise in potential of the latter causes a decrease in the current to the grid surrounding the cathode. In the case of a triode with highly positive grid and zero or low potential plate,[24] the resulting negative transconductance is not very useful because of the power consumed in the plate control element. A much improved arrangement is found in the ordinary space-charge-grid tube in which a negatively biased second grid and positively biased anode are substituted for the positive plate of the triode. Schottky pointed out[16] that in such a tube a rising second grid potential reduced the inner grid current and that the negative transconductance may be quite high. However, the resistance of the first grid, (the anode of the negative transconductance) is ordinarily much too low unless the cathode temperature is reduced. This, however, leads to lack of reliability. In the Bibliography[25–30] there will be found a few of the applications of the tetrode as a negative transconductance type of negative resistance. A great improvement is obtained in eliminating the necessity for reduced filament temperature by the addition of one or more grids between the cathode and the output electrode.[29,31,32] In a later section, the use of the 57 tube in this fashion will be found discussed in some detail.

3. *Secondary Emission Type*

In a tetrode used as a dynatron in the region in which the dynode current is negative, the first grid-to-dynode transconductance is negative. In this case the anode resistance of the tube is negative whereas in both of the previous types it is positive, in general. The negative transconductance is not ordinarily great unless the secondary emissivity is high, but the effect has been used to augment the dynode negative resistance in a voltage controlled circuit by coupling the first grid to the dynode.[7,33,34] The method of achieving negative transconductance in this case has similar disadvantages to other devices making use of secondary emission.

4. *Electron Transit-Time Methods*

When electron transit time is considered, it is found that in certain ultra-high-frequency bands the tube characteristic becomes inverted.[35] Operation in such a band is equivalent to a reversal in sign of the transconductance (also a reduction in its magnitude) and an ordinary posi-

tive transconductance tube used in such a band may produce negative resistance by direct coupling of output to input. Too little is known of these effects to permit further discussion.

5. *Magnetically Controlled Electron Tubes*

The magnetron (not the split-anode type) may be used in direct coupled circuits to achieve negative resistance of either the current or the voltage controlled type.[3] The self-inductance as well as the inconvenience of the field winding makes this means of getting negative resistance a poor one.

Reverse Phase Coupled Devices

In this group are included the well-known arrangements by which negative resistance is obtained from an amplifying system by coupling of output to input through a phase-reversing circuit or system. Most of this group is too well-known to require even a brief discussion but for illustration two of the most important classes are listed.

1. *Conventional Vacuum Tube With Feedback*

2. *Kallirotron and Multistage Vacuum Tube Amplifier*

In order to obtain negative resistance, the output is coupled to the input. The reversal of phase is accomplished either by the coupling means or by one of the tubes in the amplifier.

A Few Applications of Negative Resistance and Negative Transconductance

A negative resistance may be connected with other circuit elements for the excitation of oscillations. Such circuits have the advantage of simplicity and good frequency stability. A voltage controlled negative resistance may be used in parallel with an antiresonant circuit to produce oscillation at a frequency near resonance and a current controlled device may be used with a series resonant circuit for the same purpose. Relaxation oscillations[5] are possible with any negative resistance provided the external circuit is suitable but comparatively little is to be found in the literature using voltage controlled devices for this purpose although there are frequent descriptions using the glow discharge and other current controlled arrangements.

As a circuit element in the stable as contrasted with the oscillatory state, a negative resistance has many uses. The neutralization of positive resistance is quite feasible theoretically,[8,36] but can only be accomplished to a limited extent experimentally because of instability. A class of circuits utilizing negative resistance is that including the

networks which resemble negative capacitance or negative inductance or combinations of them.[37,38,39,40] Very little experimental work has been done with the networks possible although the circuits offer very attractive features. The realization of impedances which vary as the nth power of the applied frequency (where n is any positive or negative integer) is possible theoretically. A recently disclosed application of negative resistance to tuning is of some interest.[41]

For measurement work at both low and high frequencies, negative resistance arrangements have been found quite useful. The start of oscillation, which implies zero circuit resistance, may be easily and accurately detected so that if the value of the negative resistance is known at this point the external circuit resistance is known.[42–45] Negative resistances associated with calibrated circuits in an oscillating state have frequently been used as heterodyne-frequency meters as well as signal generators. The simplicity and constancy of frequency of the circuits make them advantageous.[46,47]

In addition to the above applications, some arrangements, particularly the negative transconductance tubes, have other uses than as a negative resistance. For example, a negative transconductance tube may be used as an amplifier in exactly the same way as the usual positive transconductance tube. The input impedance of a negative transconductance tube resulting from feed-back through the grid-plate capacitance is quite different from that of a positive transconductance tube giving equivalent amplification. In the usual case a resistive anode load gives rise to a higher equivalent input capacitance, whereas with a negative transconductance tube, the equivalent input capacitance is reduced by feed-back. A reversal of sign of the equivalent input conductance is also found in the negative transconductance tube as compared with the positive, when reactive anode loads are used. Another amplifier application possible with negative transconductance tubes of certain kinds is one in which the output is taken from two anodes in opposition fashion, a common control element having positive transconductance to one anode and negative transconductance to the other. The retarding-field type is an example; references 27, 31, 48–54 describe the application. Such a circuit, if both positive and negative transconductance parts have similar characteristics, permits the grid-plate capacitance feed-back to be eliminated. The negative transconductance tube is, of course, suitable as a modulator or detector when operated over nonlinear portions. A large number of circuit arrangements using regeneration with simple resistances are also possible with negative transconductance tubes, and especially so with the retarding-field tubes.[31,51,55]

The Use of a 57 as a Retarding Field Negative Transconductance Tube

In investigating a number of commercially available tubes for their possibilities as negative resistance devices, it was found that one of the most useful and promising arrangements was a negative transconductance one using a regular 57 tube with a negatively biased third grid (suppressor) to control the current to the second grid (screen). The transconductance between the third grid (control element) and the second grid (output anode) is negative, as would be expected. The control element, being negatively biased, draws no current and the screen or output grid has an advantageously high resistance because of the limiting action of the first grid on the cathode current. The first grid functions primarily in the indicated limiting action; however, it serves an additional useful purpose in that, by variation of its potential, the negative transconductance of the tube may be varied. The plate electrode plays no part in the circuit but must be connected to a positive potential in order to collect the electrons which pass the third grid.

A set of operating voltages for a typical 57, which does not exceed safe limits, together with the usual parameters for the tube considered as a negative-transconductance device is as follows:

Heater voltage	= 2.5 volts	Heater current	= 1.0 amp.
First grid voltage	= 0 volts	....	
Second grid voltage	= + 100 volts	Second grid current	= 3.7 milliamp.
Third grid voltage	= − 10 volts	....	
Plate voltage	= +22.5 volts	Plate current	= 3.3 milliamp.

$$\frac{\partial I_{c_2}}{\partial E_{c_3}} \doteq s_{m_{3-2}} = -310 \times 10^{-6} \text{ mhos}$$

$$\frac{1}{\dfrac{\partial I_{c_2}}{\partial E_{c_2}}} = r_{g_2} = 40{,}000 \text{ ohms}$$

$$s_{m_{3-2}} r_{g_2} = \mu_{3-2} \doteq -12.4$$

$$C_{g_2-g_3} = 3.1\ \mu\mu\text{f} \qquad C_{g_3-k} = 5.0\ \mu\mu\text{f} \qquad C_{g_3-k} = 5.7\ \mu\mu\text{f}$$

The characteristics of a 57 used in this fashion are shown by the screen-voltage vs. screen-current characteristics with various third grid voltages. Such a set of characteristics is shown for a typical 57 in Fig. 7. The dotted curves indicate the negative resistance characteris-

tic to be expected with two different operating conditions in a voltage controlled circuit using a large condenser to couple the second and third grids.

The voltage controlled negative resistance obtained under the operating conditions tabulated above is 3500 ohms, and the total cathode current is only seven milliamperes. This indicates performance superior to that of most dynatrons which at this value of cathode current have a negative resistance seldom less than 10,000 ohms. The combined

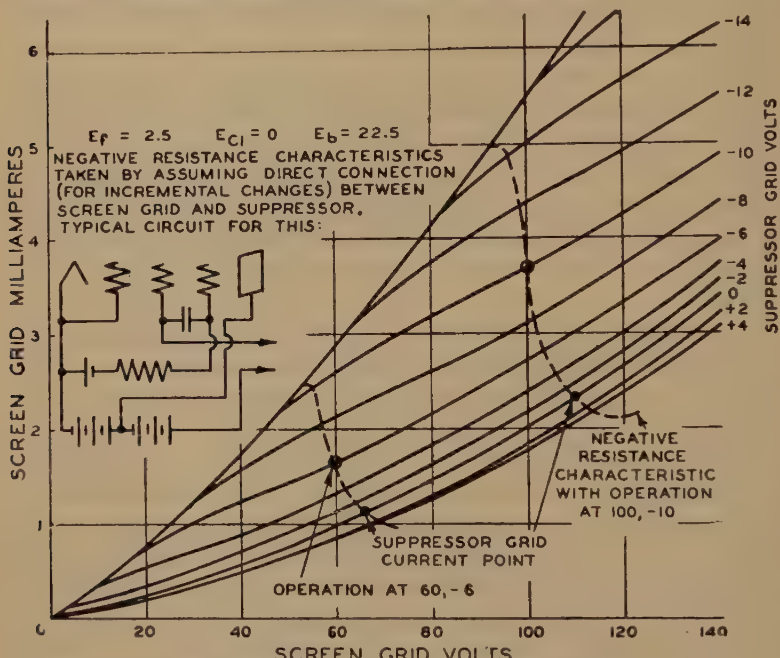
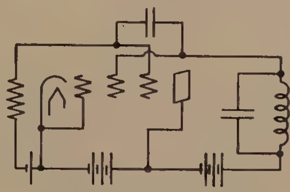


Fig. 7—Average RCA-57 used as negative transconductance tube.

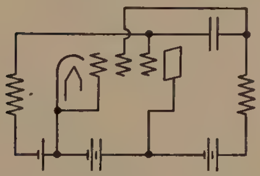
capacitance to cathode of the two grids of the 57 is roughly the same as that of the plate to cathode in a 24 tube used as a dynatron. The additional capacitance of an external coupling condenser and grid leak to ground may be made small, when necessary, so as to increase the total capacitance but slightly. Variation of first grid bias of the 57 is found to be a simple means of changing the negative resistance over as wide a range as may be desired. It is found also that variation in characteristics from tube to tube at fixed operating potentials is no greater than occurs with the normal characteristics of the 57. In addition, individual tubes are found to maintain very constant characteristics. Relatively low values of negative resistance may be obtained

with extremely low operating potentials and currents, an advantage not shared by many other devices.

Fig. 8 shows the 57 as a voltage controlled, negative resistance oscillator in conjunction with a tuned circuit. A relaxation oscillation circuit using the 57 in a current controlled circuit is shown in Fig. 8(b). It is interesting to note that a circuit similar to the latter but using a tetrode was suggested by van der Pol[5] and investigated with a 22 tube by Page and Curtis[29] who suggested that the chief limitations and drawbacks of the circuit would be overcome by a tube with three grids used in the manner of Fig. 8(b). The condition for the start of oscilla-



(a) VOLTAGE-CONTROLLED OSCILLATOR CIRCUIT



(b) CURRENT-CONTROLLED RELAXATION-OSCILLATOR CIRCUIT

Fig. 8—Oscillator circuits using a 57 tube as negative resistance.

tion as well as the frequency for infinitesimal oscillation amplitudes in a circuit of this kind has been derived above in the discussion of negative transconductance tubes in current controlled, negative resistance circuits. It should be pointed out that the oscillations in a circuit of this kind are approximately sinusoidal if operation at small amplitudes can be maintained, such as is possible by operation at the point of peak negative transconductance, $(\partial^2 I_b/\partial E_c^2=0)$, and by adjustment to the critical condition for the start of oscillation.

Acknowledgment

This study of negative resistance was conducted in the Research and Development Laboratory of the RCA Radiotron Company, Inc. I wish to acknowledge my appreciation of the contributions, too numerous to list individually, made during discussions with various members of the engineering staff. In particular, I would like to thank Mr. T. M. Shrader, under whose direction this study was made, for his encouragement and Mr. B. Salzberg for his many helpful suggestions.

Bibliography

(*Titles of articles are freely translated in some cases*)

(1) G. Crisson, "Negative impedances and the twin 21-type repeater," *Bell Sys. Tech. Jour.*, vol. 10, p. 485, (1931).

(2) H. Barkhausen, "Stability conditions of the arc light and electron tubes," *Phys. Zeit.*, vol. 27, p. 43, (1926).

(3) K. Steimel, "Stability and self-oscillation of electric circuits including devices with falling characteristics," *Zeit. für Hochfrequenz.*, vol. 36, p. 161, (1930).

(4) K. Heegner, "Self-excited oscillation produced by means of vacuum tubes in series and parallel tuned circuits," *Hochfrequenz. und Elektroakustik*, vol. 40, p. 198, (1932).

(5) B. van der Pol, "Relaxation oscillations," *Phil. Mag.*, vol. 2, p. 978, (1926).

(6) W. Reichardt, "The degeneration of sinusoidal oscillations," *Elek. Nach. Tech.*, vol. 8, p. 344, (1931).

(7) N. W. McLachlan, "On the influence of valve resistance in oscillation generators," *Wireless Eng. and Eng. Wireless*, vol. 9, p. 130, (1932).

(8) A. W. Hull, "Description of the dynatron," Proc. I.R.E., vol. 6, p. 5; February, (1918).

(9) J. Scott-Taggart, "The negatron," *Electrician*, vol. 87, p. 386, (1921), and *Radio Rev.*, vol. 2, p. 598, (1921).

(10) R. Seeliger and G. Mierdel, "Self-sustaining discharges in gases," *Handbuch der Experimentalphysik*, vol. 13, part 3, (1929).

(11) F. M. Penning, "Double valued negative characteristics in self-sustained gas discharges," *Phys. Zeit.*, vol. 33, p. 816, (1932).

(12) P. L. Copeland, "Surface conditions and stability of characteristics in screen-grid tubes," *Jour. Frank. Inst.*, vol. 216, p. 417, (1933).

(13) M. G. Scroggie, "Applications of the dynatron," *Wireless Eng. and Eng. Wireless*, vol. 10, p. 527, (1933).

(14) K. Rottgardt, "Generation of oscillation in triode," *Zeit. für tech. Phys.*, vol. 6, p. 610, (1926).

(15) E. Friedlander, "Direction phenomena through 'field decay' and relaxation oscillations in electron tubes," *Zeit. für tech. Phys.*, vol. 7, p. 481 (1926).

(16) W. Schottky, "High vacuum amplifiers," *Arch. für Elektrotech.*, vol. 8, p. 299, (1919).

(17) E. W. B. Gill, "Negative Resistance Tube," British Patent 251,701, (1925).

(18) E. Alberti, "An investigation of space-charge-grid tubes," *Elek. Nach. Tech.*, vol. 3, p. 149, (1926).

(19) E. L. Chaffee, "Theory of Thermionic Vacuum Tubes," pp. 242–260, McGraw-Hill Book Co., Inc., (1933).

(20) E. Habann, "A new generator tube," *Zeit. für Hochfrequenz.*, vol. 24, pp. 115 and 135, (1924).

(21) W. E. Benham, "Theory of the internal action of thermionic systems at moderately high frequencies," *Phil. Mag.*, vol. 6, p. 641, (1928), and vol. 11, p. 457; February, (1931) Suppl.

(22) J. Muller, "Electron oscillations in high vacuum," *Hochfrequenz. und Electroakustik*, vol. 41, p. 156, (1933).

(23) F. B. Llewellyn, "Vacuum tube electronics at ultra-high frequencies," Proc. I.R.E., vol. 21, p. 1532; November, (1933).

(24) H. E. Hollmann, "Reception of ultra-short waves with the retarding field audion," *Hochfrequenz. und Elektroakustik*, vol. 42, p. 89, (1933).

(25) K. C. Van Ryn, "The Numan's oscillator," *Eng. Wireless and Wireless Eng.*, vol. 2, p. 134, (1924).

(26) E. Mittelman, "Experiments with the Numans-Roosenstein oscillatory circuit," *Zeit. für Hochfrequenz*, vol. 28, p. 188, (1926).

(27) B. Decaux, "Employment of four-electrode valves," *L'Onde Electrique*, vol. 6, p. 1, (1927).

(28) R. H. Elsner, "The negadyne circuit," *Radio Bildfunk, Fernsehen für Alle*, vol. 8, p. 299, (1929).

(29) R. M. Page and W. F. Curtis, "The van der Pol four-electrode tube relaxation oscillation circuit," Proc. I.R.E., vol. 18, p. 1921; November, (1930).

(30) T. von Nemes, "Artificial arc circuits using two-grid valves," *Zeit. für Hochfrequenz.*, vol. 38, p. 77, (1931).

(31) J. C. Warner, U.S. Patent 1,756,893, applied for July 1, 1925; issued 1930.

(32) R. Wigand, "The hexode, a new German valve," *Radio Bildfunk, Fernsehen für Alle*, vol. 12, p. 279, (1933).

(33) F. M. Colebrook, "The dynatron oscillator," *Wireless Eng. and Eng. Wireless*, vol. 8, p. 581, (1931).

(34) W. van B. Roberts, "A reversed-current feed-back oscillator," *QST*, vol. 16, p. 32; February, (1932).

(35) H. E. Hollmann, "The excitation of oscillations through feedback with consideration of the ultra-dynamic characteristic," *Sitz. Ber. d. preusz. Ak. d. Wiss. Phys. Math. Klasse*, VI 41.S., (1933).

(36) F. M. Colebrook, "Voltage amplification with high selectivity by means of the dynatron circuit," *Wireless Eng. and Eng. Wireless*, vol. 10, p. 69, (1933).

(37) A. C. Bartlett, "Bourcherot constant-current networks and electric wave filters," *Jour. I.E.E.* (London), vol. 65, p. 373, (1927).

(38) B. van der Pol, "A new transformation in alternating-current theory with an application to the theory of audition," *Phil. Mag.*, vol. 43, p. 477, (1929), and Proc. I.R.E., vol. 18, p. 221; February, (1930).

(39) L. C. Verman, "Negative circuit constants," Proc. I.R.E., vol. 19, p. 676; April, (1931).

(40) L. C. Verman, "Transients in negative constant series circuits," *Jour. Indian Inst. Sci.*, vol. 55 B, part III, p. 33.

(41) S. Cabot, "Resistance tuning," Proc. I.R.E., vol. 22, p. 709; June, (1934).

(42) H. Pauli, "Measurement of electric effective resistance by the help of negative resistance," *Zeit. für tech. Phys.*, vol. 10, p. 592, (1929).

(43) H. Iinuma, "A method of measuring the radio-frequency resistance of an oscillatory circuit," Proc. I.R.E., vol. 18, p. 537; March, (1930), and vol. 19, p. 467; March, (1931).

(44) H. Fruhauf, "Measurement of loss resistance of high-frequency oscillating circuits by the dynatron," *Zeit. für Hochfrequenz*, vol. 87, p. 229, (1931).

(45) A. T. Green, "High-frequency resistance of coils," *Eng. Wireless and Wireless Eng.*, vol. 8, p. 183, (1931).

(46) G. Grammer, "The dynatron frequency meter," *QST*, vol. 14, p. 9; October (1930).

(47) W. H. F. Griffiths, "The simplification of accurate measurement of radio frequency," *Wireless Eng. and Eng. Wireless*," vol. 10, pp. 239 and 299, (1933).

(48) H. J. van der Bijl, U.S. Patent 1,479,779, (1924).

(49) L. Muller, "Compensation of the anode reaction effect," *Arch. für Elektrotech.*, vol. 16, p. 251, (1926).

(50) R. Barthelemy, "Resonance amplification with two-grid valves," *L'Onde Electrique*, vol. 6, p. 152, (1927).

(51) J. C. Warner, "Some characteristics and applications of four-electrode tubes," Proc. I.R.E., vol. 16, p. 424; April, (1928).

(52) M. von Ardenne and W. Stoff, "The compensation of undesired capacitances and their effects," *Zeit. für Hochfrequenz.*, vol. 31, p. 122, (1928).

(53) Anonymous, "The isodyne circuit with double grid valves," *Wireless World*, vol. 24, p. 465, (1929).

(54) C. Krulisz, "A pseudo-symmetrical receiving circuit using one four-electrode valve," *Wiadomosci i Prace Inst. Radjotech.*, (Warsaw), vol. 2, p. 209, (1931).

(55) W. Buhk, "The behavior of a space-charge-grid tube when using regeneration in the space-charge-grid circuit," *Hochfrequenz. und Elektroakustik.* vol. 42, p. 152, (1933).

Proceedings of the Institute of Radio Engineers
Volume 23, Number 10 *October, 1935*

COMPARATIVE ANALYSIS OF WATER-COOLED TUBES AS CLASS B AUDIO AMPLIFIERS*

By

I. E. MOUROMTSEFF AND H. N. KOZANOWSKI

(Westinghouse Electric and Manufacturing Company, East Pittsburgh, Pennsylvania)

Summary—*Performance of vacuum tubes used as class B audio amplifiers has been studied with particular stress on the influence of the amplification factor, μ, on the behavior of tubes. Four types of water-cooled tubes of completely identical design differing only in μ have been studied. The influences of the load resistance, operating plate voltage, and no-signal current, or bias, have been graphically pre-calculated and compared. The limitations of water-cooling on this class of tube operation are clearly demonstrated. The methods used for this particular study are applicable to a variety of similar problems.*

THE purpose of this investigation is the analysis from tube characteristic charts of the performance of vacuum tubes used as class B audio amplifiers. Particular attention is given to the influence of the amplification factor, μ, on the behavior of various tubes. Such an investigation becomes more interesting at the present time when high efficiency plate modulators have established their place in radio practice.[1] These modulators are merely last-stage audio amplifiers supplying power at audio frequencies to the plate circuit of the radio-frequency output stage of a radio transmitter.

Three types of standard water-cooled tubes, the UV-863, UV-207, UV-848, and a special type WL-419A made for this purpose by the Westinghouse Lamp Company were subjected to this study. The reason for this choice of tubes is that all four types have identical structure in every respect, except for the number of turns of grid winding, making the amplification factor different with each type of tube; namely,

with	UV-863	$\mu=50$
	UV-207	20
	UV-848	8
	WL-419A	3.6

Thus the chosen tubes cover the entire range of all feasible voltage factors usually encountered in high power transmitting tube practice.

* Decimal classification: R330×R355.7. Original manuscript received by the Institute, April 12, 1934; revised manuscript received by the Institute, December 10, 1934. Presented before Ninth Annual Convention, Philadelphia, Pa., May 28, 1934.

[1] J. A. Hutcheson, "Application of transformer coupled modulators," PROC. I.R.E., vol. 21, pp. 944–958; July, (1933).

The Problem Outlined

The postulated general problem directly implies the following questions: Which of the four types of tubes gives the least distortion for a given output? Or, conversely, which tube gives the greatest output without exceeding a prescribed distortion limit?

In connection with the outlined problem it was necessary to develop certain graphical methods which facilitate the systematic investigation of tube performance directly from conventional charts of tube characteristics. These methods can usefully be applied to any similar study.

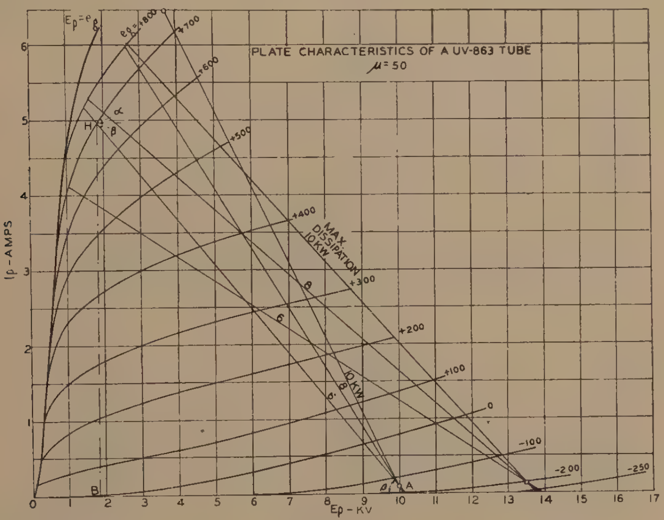


Fig. 1

Dynamic Characteristics

The first step in this and similar work is the preparation of complete charts of plate characteristics for the tubes to be investigated, if such charts are not already available. In plotting the charts of Figs. 1 to 4, the usual static method was used to record the curves in the negative grid-voltage region; for positive grid potentials the oscillographic method was used.[2] It must be noted that the static characteristics shown in these charts belong to individual tubes available at the time of this study. Therefore, the absolute quantitative results may depart somewhat from true average figures. Yet, this fact can hardly affect the general picture of the *comparative* properties of the investigated tube types.

[2] H. N. Kozanowski and I. E. Mouromtseff, "Vacuum tube characteristics in the positive grid region by an oscillographic method," Proc. I.R.E., vol. 21, pp. 1082–1097; August, (1933).

A systematic study of tubes can be achieved by comparison of the results obtained for each tube under identical operating conditions. According to definition, a class B audio amplifier is operated at, or

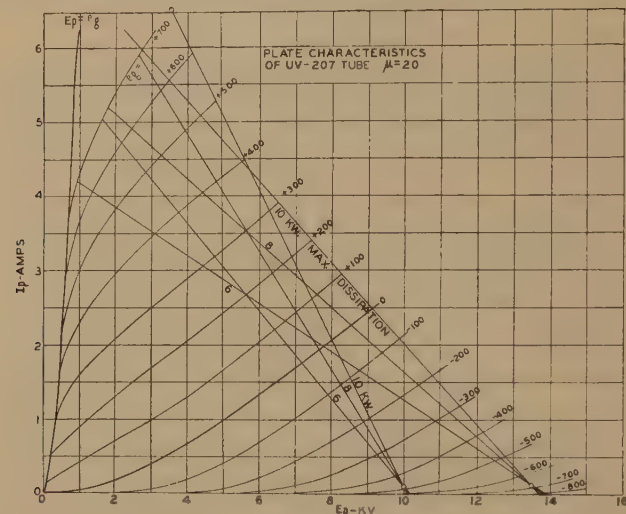


Fig. 2

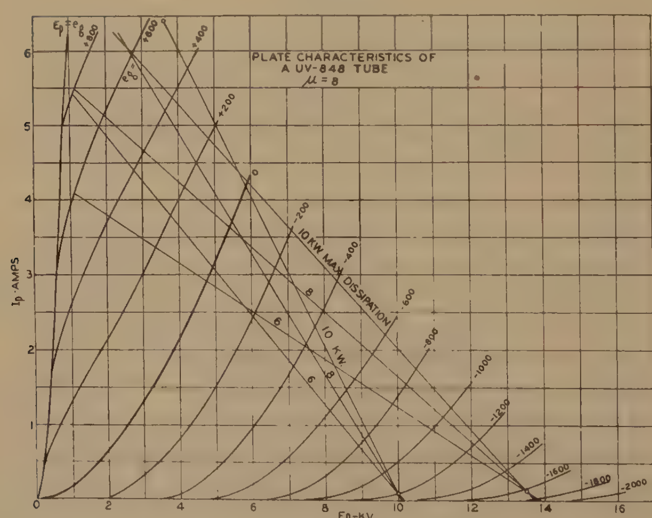


Fig. 3

nearly at, the cutoff bias for the assumed plate voltage, so that only a small residual or *no-signal* plate current flows when no excitation is applied to the grid. The alternating output current flows during essen-

tially one half of the cycle, when the grid excitation voltage swings towards its positive values. The magnitude of the output plate current depends on the value of the output or load resistance connected between plate and filament of the tube. Thus operating conditions are fully determined by:

(1) Operating plate voltage, E_p

(2) Load resistance, R_L

(3) No-signal current, I_o, determined by the grid bias, e_c.

The influence of these three factors can be studied separately. By correlating the results of such procedure a final conclusion regarding an optimum value of μ can be drawn.

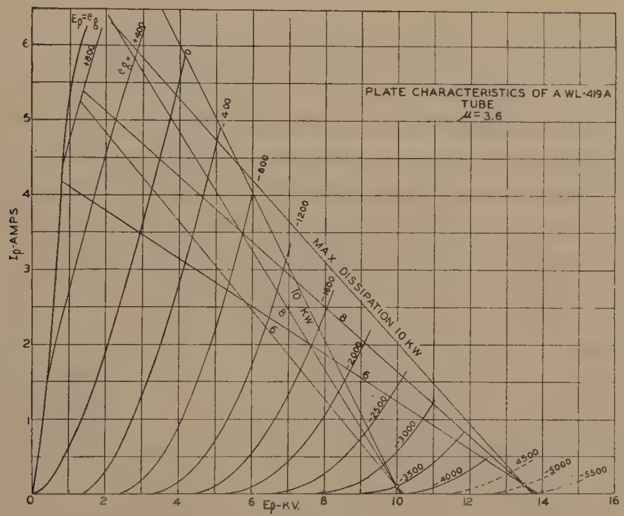


Fig. 4

For operating voltages two values are chosen: 13,500 volts, which is considered the highest practicable voltage for this class of operation, and 10,000 volts, which is the voltage quite commonly used in practice on water-cooled tubes of the types under investigation.

The no-signal current, I_o, was chosen in accordance with existing practice. It is 100 milliamperes with 10,000 volts, and 125 milliamperes with 13,500 volts on the plate, common to all tubes.

The choice of load characteristics was governed by the assumed maximum plate dissipation, P_h. Two factors allow for a rapid plotting of load characteristics for any chosen value of P_h.

(1) In the plate current-voltage charts the load characteristics for audio output are straight lines conventionally plotted through the

chosen operating point (E_p, I_o). This is because the load in the case of a modulator can be considered as a pure *ohmic resistance.*

(2) The largest anode dissipation in the investigated mode of operation with any load resistance invariably corresponds to fifty per cent efficiency. This, neglecting distortion, always occurs for a plate voltage swing of $(2/\pi)E_p$, leaving the voltage across the tube, E_{min} equal to 0.364 E_p. The current ordinate, I, for this condition will be found from the expression

$$P_h = P_o = 1/4 \times 0.636 E_p I \tag{1}$$

which states the condition of equality of output power and plate loss. By these two points the load line is completely defined for any assumed value of P_h.

Using this procedure, three load characteristics for each operating plate voltage are plotted in the chart of each tube under the assumption of 10, 8, and 6 kilowatts dissipation, respectively. The largest value, 10 kilowatts, is the one conventionally accepted as the maximum dissipation allowable for standard water-cooled tubes. The other two have been chosen in the anticipation that better performance can be secured with less steep dynamic characteristics, or in other words, with higher values of load resistance, R_L, given by the inverted slope of the dynamic characteristic. The actual values chosen are shown in Table I.

TABLE I

E_p	10,000 volts	13,500 volts
P_h = 10 kw	R_L = 1000 ohms	R_L = 1850 ohms
8	1260	2300
6	1680	3130

Power output, P_o, and distortion δ per cent, were studied along each load line up to the highest practicable value of grid excitation. Actually, all plotted load lines or *modulation paths* are terminated at, or in the vicinity of, $e_p = +700$ volts, beyond which the distortion rapidly increases to quite a prohibitive degree due to the approach to the diode line.

Transconductance Modulation Curves

In order to judge distortion for various grid excitations, one must assume the grid voltage as independent variable, because, according to definition, in an ideal class B amplifier the output current and voltage must be proportional to the input or grid excitation voltage. Therefore, from each straight-line modulation path drawn in the plate charts a transconductance modulation characteristic having grid voltage as independent variable is plotted point by point. Such curves are shown

in Figs. 5 and 6 for the UV-863 and WL-419A tubes as typical examples, having, respectively, the highest and lowest values of μ. In each drawing three curves are given in accordance with the three chosen plate dissipations of 6, 8, and 10 kilowatts.

For graphical analysis of tube performance with any grid excitation, $e_o = e_{g_{max}} + e_c$, the extreme end of the utilized portion of the operating characteristic is connected by a straight line with the operating point or zero-signal end as shown in Fig. 5. Then the ordinate differences between the characteristic curve and the spanning chord are de-

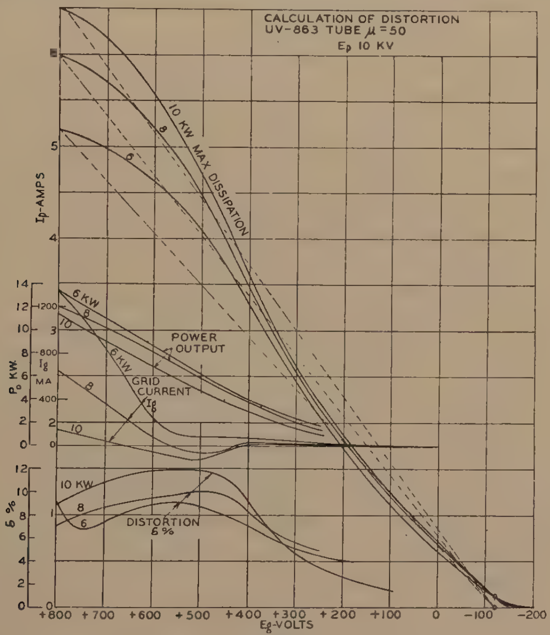


Fig. 5

termined for intermediate grid voltages equal to 0.309, 0.5, 0.707, 0.809, and 0.866 of the assumed grid voltage amplitude, e_o. From these data the amplitude of the 3rd, 5th, 7th, 9th, and 11th harmonics, and the true amplitude of the fundamental frequency can be calculated in the manner outlined in an article on ordinate-difference harmonic analysis.[3]

The procedure is repeated for four or five points along each characteristic, corresponding to four or five different grid excitations. This allows the plotting and determination of smooth curves showing the

[3] I. E. Mouromtseff and H. N. Kozanowski, "'Short-cut' method for calculation of harmonic distortion in wave modulation," Proc. I.R.E., vol. 22, pp. 1090–1102; September, (1934).

variation of distortion along the entire modulation path as function of the grid excitation or *depth of modulation*. Power output for each point can also be calculated easily. Then, it remains a matter of the designer's choice to select for actual operation as large a portion of a particular modulation characteristic as he deems suitable. The power output and

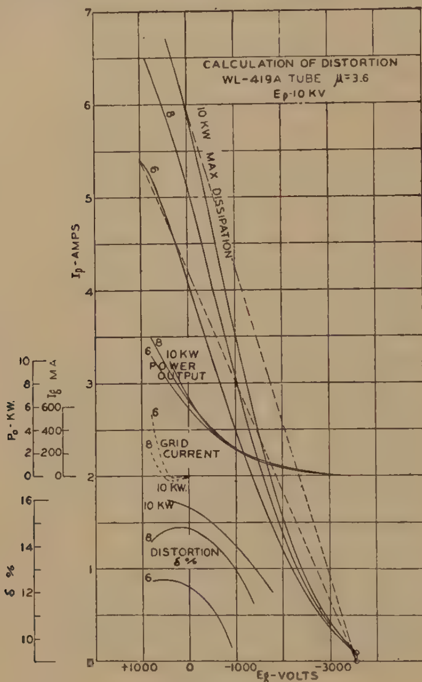


Fig. 6

distortion curves are plotted in both transconductance charts, together with the instantaneous value of the grid current along any path. The following table was compiled from the calculated curves of distortion, giving percentages of distortion for all three load resistances or maximum plate dissipations at two operating plate voltages.

TABLE II
DISTORTION WITH CONVENTIONAL (NOT OPTIMUM) BIAS

		$P_o=10$ kw								
		$E_p=10{,}000$ volts						$E_p=13{,}500$ volts		
		100% Modulation			75% Modulation			75 to 100% Modulation		
Tube Type	μ	$P_h=6$	8	10 kw	6	8	10 kw	6	8	10 kw
UV-863	50	7.2	8.3	9.8	7.8	9.2	11.3	8.0	10.5	11.2
UV-207	20	6.0	7.8	9.0	8.8	8.5	9.4	6.0	7.6	8.2
UV-848	8	9.3	10.8	10.8	10.4	11.4	12.2	5.7	7.0	9.6
WL-419A	3.6	12.5	12.4	16.0	12.5	14.8	16.0	11.0	13.0	14.0

Per cent Distortion

Results of Graphical Investigation of Conventional Operating Conditions

A direct inspection of Figs. 5 and 6 and of similar curves for all operating conditions and of the above table reveals that:

(1) For a given grid excitation the power output, P_o, for each tube varies but little for each of the *confluent* load lines passing through the same operating point. The departure is generally greater with 13,500 volts on the plate than with 10,000 volts.

(2) The six-kilowatt line, the lowest of a confluent set, gives somewhat greater output than the eight- or ten-kilowatt lines for the same grid excitation with tubes having high μ (UV-863). This is explained by the fact that for maximum output the angle between the load line HA, as shown in Fig. 1, and the static curve e_g = constant, on which the load line is terminated, must be such that a horizontal line through its vertex bisects this angle. This gives $\alpha = \beta$ in the figure. Hence, with the steeper static characteristics of low-μ tubes the dynamic load line giving a maximum output for a definite maximum grid voltage must have a greater slope than with high-μ tubes. Thus, with the UV-863 tube and a maximum grid voltage of, say, +700 volts, the maximum output happens to be in the vicinity of the six-kilowatt load line. With low-μ tubes, such as the UV-848, the maximum output for any grid voltage corresponds to a much steeper dynamic characteristic and a larger output power than along the six-kilowatt line. In fact, such a characteristic is so steep that it cannot be utilized because of too great an associated plate dissipation. The statement regarding the optimum position of a dynamic load line can easily be proved analytically; this is done in an addendum.

(3) Distortion throughout is lower for low dissipation lines with all tubes for both operating voltages.

(4) In order to secure reasonably high output from the investigated tubes under the specified operating conditions one has to tolerate distortions from about six to sixteen per cent. Assuming that the highest permissible distortion for a transmitter can be only ten per cent[4] the calculated distortion seems to be too high for practical use under conventional conditions. Yet, in combination with preceding stages of amplification it is feasible to secure more favorable over-all results as the distortions introduced in the various stages may be partly cancelled out.

(5) For each individual transconductance modulation curve, concave or convex in form, and lying entirely on one side of the spanning chord, the distortion increases with grid swing. With more or less symmetrical S-shaped curves, the maximum distortion is located near the

point where the chord is tangent to the upper bend of the curve. In the vicinity of the diode line the distortion rises abruptly due to the crowding of the static curves.

(6) With 10,000 volts on the plate distortion along each modulation path is, in general, smaller for tubes with higher μ (UV-863 and UV-207). At 13,500 volts, tubes having both extremly high μ (UV-863) and extremely low μ (WL-419A) produce greater distortion than tubes of medium μ such as the UV-207 and UV-848. In addition, one may

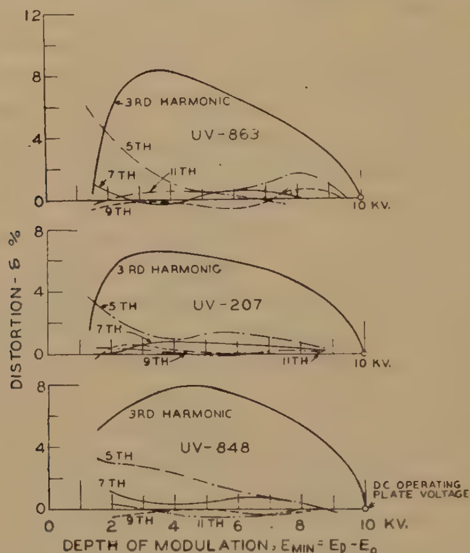


Fig. 7—Distribution of harmonic components.
$E_p = 10{,}000$ volts
$R_L = 1680$ ohms
$I_0 = 0.100$ ampere

note that the fidelity of the high- and low-μ tubes decreases more rapidly with increasing slope of the load line than with the UV-207 tube having a μ of medium value.

The values of distortion plotted in Figs. 5 and 6 and given in the preceding table represent the *arithmetical sum*[4] of the odd harmonics up to the 11th order. Even harmonics are absent because of the symmetry of the push-pull arrangement. The percentage of each individual component is calculated as the ratio of its amplitude to the *true* amplitude of the fundamental frequency. The distribution of harmonic components of the distortion along the six-kilowatt modulation path with a no-signal current of 100 milliamperes for three standard tubes is

[4] Federal Radio Commission, "Rules and Regulations," Sec. 103 and 139.

shown graphically in Fig. 7. It will be noted that with high modulation, when the tube swings well into the upper portion of an S-shaped transconductance curve, the 5th harmonic is predominant, while with medium modulation the 3rd becomes more pronounced. The 9th and 11th components together usually contribute from zero to one fifth of the total distortion, and, therefore, their calculation could have been omitted entirely in the majority of the cases studied. Yet, with the harmonic analysis method used, they serve as a very good check on the correctness of the calculation of the other components just because they must come out smaller than components of lower orders.

We must take notice of one important detail in connection with the graphical method used for calculating distortion from the transconductance modulation characteristics: The chords spanning each curve are drawn *not* through the actual operating point on the curve (Fig. 5) given by its coördinates e_c and I_o, *but through a point on the zero current axis directly below the operating point.* By this procedure account is taken of the concurrent effect of the "companion" tube on the other side of the push-pull arrangement. This will be discussed in detail in a later section.

Résumé of Results Under Conventional Operating Conditions

The data obtained from the preceding study show rather definitely that under what one may term conventional operating conditions none of the four types of tubes in question gives highly satisfactory results *per se* in class B audio service. This is particularly true with the nonstandard WL-419A tube, having the unusually low μ of 3.6. With it the distortion is never below eleven per cent if a reasonably high audio output of say seven to ten kilowatts is required. In fact, under certain conditions the distortion even goes up to sixteen per cent.

Yet, if one is to make a choice from the three standard tubes on the basis of distortion with conventional operating conditions the UV-863 tube with a μ of 50 will have preference for a low voltage of 10,000 volts, while the UV-848 having a μ of 8, may have application at the higher operating value, 13,500 volts. However, the UV-207 tube with a medium μ equal to 20 appears from the analysis to be more universal, giving favorably comparable results with either operating voltage.

With this conclusion the investigation might have been considered closed. However, a study of three additional factors has been found essential for complete knowledge of the scope of application of the investigated tubes in class B audio amplifier service. These factors are brought to the foreground by the following quite natural questions:

(1) What is the effect on the performance of tubes having various values of μ, of an arbitrary up-and-down shift of the operating point in the plate current-voltage chart. In other words, what changes in tube operation are brought about by an arbitrary change of the no-signal current, I_o?

(2) What restriction is imposed by the water-cooling of the plate on the choice of operating conditions?

(3) How do the various tubes differ in their necessary grid excitations?

Before going into the discussion of these specific questions, it becomes necessary to outline certain details in the graphical method applied to their analysis.

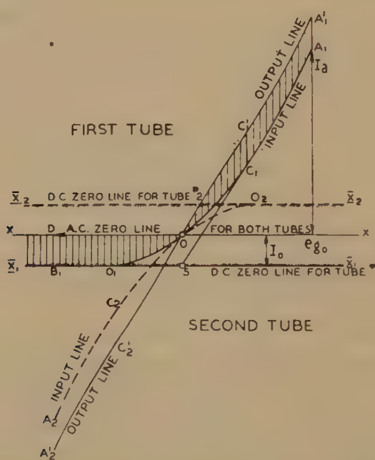


Fig. 8—Modulation characteristic for two tubes in push-pull.

Method of Graphical Treatment of Transconductance Modulation Curves

It was previously pointed out that the chord spanning a transconductance modulation curve used in the calculation of distortion must be plotted through the point on the zero current axis directly below the chosen operating point and not through the operating point itself. This procedure is justified in the following discussion:

A typical transconductance modulation curve, $A_1C_1OO_1$ for a modulator tube is shown in Fig. 8. An identical curve $A_2C_2OO_2$ is inverted and drawn in the proper relation to the first curve; it represents the modulation characteristic of the *companion* tube, both tubes working in push-pull. The common operating point, O, indicates a no-signal current, I_o. The direct-current zero lines, X_1X_1 and X_2X_2, for both tubes,

and the common alternating-current zero axis XX are shown in the drawing in their proper relation.

One can readily see that the negative alternating current during the negative half cycle of grid voltage for the first tube will be added in the output transformer to the positive current of the second tube, and vice versa. The true output curve is then given by $A_1'C_1'OC_2'A_2'$. An exactly similar output curve can be obtained from the transconductance chart of a single tube by symmetrical addition of the ordinates of the shaded area, OO_1B_1D to the positive portion OC_1A_1, of the original curve. In practice, in order to avoid replotting of the entire modulation curve, one can imagine that the new curve $OC_1'A_1'$ is simply dropped by the amount I_o. Then the portion $C_1'A_1'$ will coincide with C_1A_1, while the part C_1S can be obtained by symmetrical subtraction of the direct-current ordinates of the tailed portion of the curve from the ordinates of the original transconductance curve to the right side of the operating point.

Thus, taking into account the effect of the negative half cycle, one must replace the original or *input* dynamic characteristic $O_1OC_1A_1$ by the *output* curve SC_1A_1 passing through the point S on the direct-current zero axis, directly below the chosen operating point. This output curve rather than the original input characteristic must be considered in the computing of average current, plate loss, and efficiency.

It will be noted that in the calculation of the average plate current from the component harmonics as outlined in the "short-cut" method[3] one commits an error caused by the difference between the *output* and *input* curves represented by the area SOO_1 in Fig. 8. However, this error is quite insignificant with high modulation. It can affect the calculations only at low signal strengths, in which case suitable corrections can easily be made.

Note:

In Figs. 5 and 6 the lower portions of the input curves, corresponding to the part SC shown in Fig. 8, are omitted for the sake of simplicity. They are not sufficiently long in these cases to influence the results of calculation appreciably. In further discussion, in Fig. 13 they will be shown.

Exact Method of Plotting Load Characteristics in the Plate-Current-Voltage Charts

The load or dynamic characteristics drawn in the plate-current-voltage charts of Figs. 1 to 4 are the basis for the point-by-point plotting of transconducatance modulation curves. If the no-signal plate current is zero, there is no doubt as to how a dynamic characteristic

of a class B modulator is to be plotted in the plate chart. It is a straight line passing through A_o, the operating point, and making the vertical axis an angle β such that $\tan \beta = R_L$ ohms, as indicated in Fig. 9. Here R_L is the load resistance across the plate and filament of a *single* tube. The output wave or time curve of each tube is then represented, disregarding distortion, by one half of a sine wave. In a practical circuit of a push-pull arrangement, such as is shown in Fig. 10, each of the

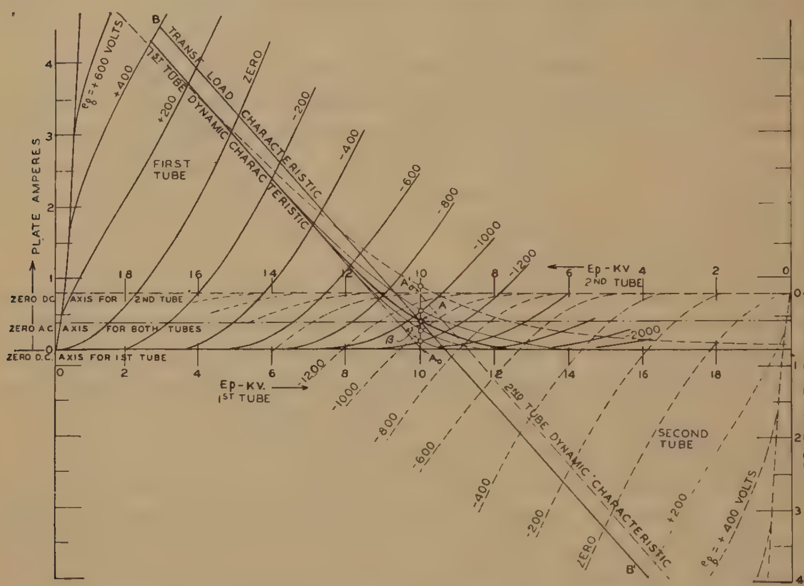


Fig. 9—UV-848 tube as class B audio amplifier. Dynamic characteristics for various values of no-signal current.

two companion tubes alternately delivers its output power *into the same physical resistance*, R_L ohms, located across the secondary of an audio transformer, the primary of which is connected between the plates of the two tubes. One may note that for modulation purposes the transformer ratio, n, between the secondary and each half of the primary winding must be made nearly unity if one intends to use the same direct-current plate voltage source for both modulator and oscillator. For 100 per cent modulation it depends exactly on the ratio of the alternating voltage amplitude of the modulator tube to the direct voltage used on the radio-frequency output stage to be modulated.

In practical operation, to avoid excessive distortion, one always chooses the operating point so that the no-signal plate current *is not zero*, for example, as shown by point A in Fig. 9. This choice, however, makes the plotting of the actual dynamic characteristics somewhat

more involved, as now power is actually delivered to the load during both half cycles in such a way that when one tube in its turn supplies the main portion of the total output, the share of the other tube is determined by the value of the no-signal current. This action can be made clear by a graphical correlation of the charts of the two companion tubes given in the same figure. Here the horizontal line passing through the mutual operating point A, represents the alternating-current axis common to both tubes. The general shape of the corresponding output waves for both tubes is shown in Fig. 11. From this, it is evident that the output wave for each tube is highly assymmetrical. During the positive half cycle the time curve is nearly sinusoidal, while on the negative half cycle it is greatly flattened. The amplitude of this

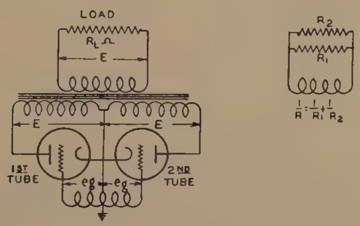


Fig. 10—Class B modulation scheme.

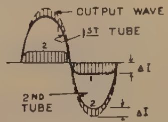


Fig. 11—Output wave.

negative half wave depends directly on the value of no-signal current chosen for operation.

The alternating currents of the two tubes are added in the secondary of the transformer in such a manner that the current flow through the load is, at every instant, equal to the sum of the two component currents. It is important to realize, further, that due to the complete symmetry and close magnetic coupling between the two halves of a push-pull arrangement, the absolute alternating values of the plate and grid voltages for both tubes are the same at any instant. From these statements it is clear that the common load, R_L, on the secondary side of the transformer is at any moment divided between the tubes so that the relation

$$1/R_L = 1/R_1 + 1/R_2. \tag{2}$$

is always preserved. R_1 and R_2 are the loads reflected on tubes 1 and 2; they vary continuously during an audio cycle so that the above relation is always satisfied. The particular case

$$R_1 = R_2 = R_L/2 \tag{3}$$

holds in general only for the instant when both tubes are sweeping through the operating point, A.

With these necessary conditions in mind one can arrive at the following rules for plotting exact dynamic characteristics for class B audio modulators in the plate charts.

(1) *The transformer load characteristic, BAB'*, referred to the secondary of the transformer is a straight line passing through the operating point A, and making an angle $\tan^{-1} R_L$ with the vertical axis. This is correct irrespective of the actual shape of the output wave. The transformer characteristic has no direct relation to the grid voltages which it crosses in the chart, and is, in general, not to be used for computing distortion.

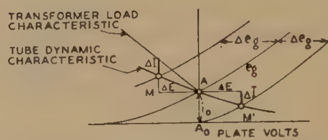


Fig. 12—Plotting of the tube dynamic characteristic.

(2) *A dynamic characteristic for each tube* can be plotted in the following way: One assumes a definite grid voltage increment Δe_g, on both sides of the operating point A. Then one finds two points M and M' on two static characteristics, $e_c + \Delta e_g$ and $e_c - \Delta e_g$, symmetrical with respect to the operating grid bias, which must satisfy the two following requirements: (a) They must correspond to equal alternating output voltage increments, ΔE, to the right and to the left of the operating point, A. (b) They must correspond to equal current increments ΔI. For the positive half cycle ΔI is the vertical distance from the chosen point, M, to the transformer load characteristic, AC. On the negative half cycle ΔI is the vertical distance between the point M' and the alternating-current zero axis or horizontal line passing through the operating point. Then M and M' belong to the dynamic characteristic of *a single tube*. The complete curve constructed by this method is used for plotting the corresponding transconductance characteristic from which the distortion is calculated in the manner described in previous sections.

Using the method outlined one finds that the main portion of the dynamic characteristic of a tube is generally a straight line dropped below the transformer load characteristic by the amount of the no-signal current, I_o. This portion corresponds to $R_2 = \infty$, and, hence, to $R_L = R_1$. In the vicinity of the operating point such a dynamic characteristic becomes a curved line tangent to the straight portion and to the *zero direct-current* axis. This curved portion is of no importance in calculations when the no-signal current is low, but the higher is the

no-signal current, the more pronounced is the influence of the exact plotting of the dynamic curve on the accuracy of the final results. For this reason in the following sections, where higher no-signal currents are considered, the method just described is actually applied to the plotting of all dynamic characteristics. Examples of dynamic and transconductance characteristics plotted by the outlined method are given in Figs. 9 and 13 for the UV-848 tube. In Fig. 9, for the sake of clearness of the drawing, the curved portions of the dynamic lines of the second tubes are not shown.

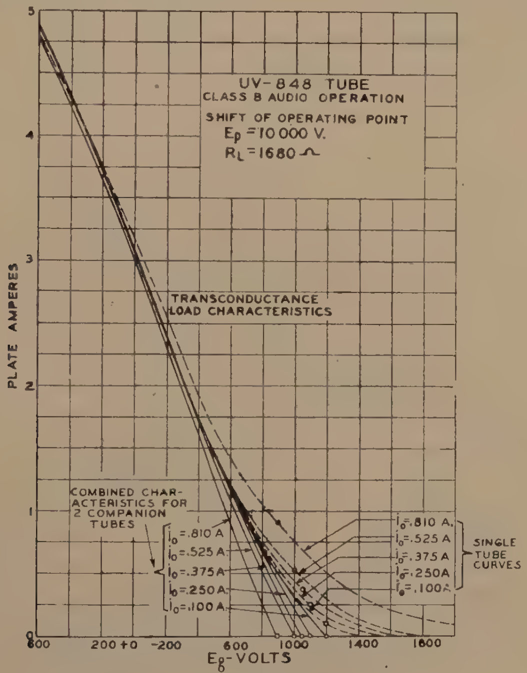


Fig. 13

One may note that with certain sets of operating conditions the tailed-off portion of a dynamic characteristic does not quite reach the direct-current zero line of the tube chart on the negative half cycle. This occurs with low-μ tubes at higher values of no-signal current, particularly if the dynamic load line is not so steep. However, this does not affect the described graphical procedure, except for the fact that the straight portion of the dynamic characteristic no longer coincides with the line passing through the point A_o; it lies above it by an amount $I_o - I_a$, the difference between the no-signal current and the current amplitude in the negative half cycle. An example of such a curve is to be seen in the same Figs. 9 and 13, shown in broken lines.

Shift of Operating Point with a Constant Plate Voltage

In order to find what comparative results in the reduction of distortion can be achieved with tubes having different values of μ as the no-signal current is changed, a calculation of operating data was carried out for the same four types of water-cooled tubes. The operating voltage throughout this investigation was taken as 10,000 volts and the load resistance across the secondary of the output transformer with a unity turns-ratio is equal to 1680 ohms, corresponding to the six-kilowatt plate loss line.

For each tube several operating points were chosen in the plate-current-voltage chart on the vertical line $E_p = 10{,}000$ volts for no-signal current values ranging from 100 to 800 milliamperes. The exact location of each operating point was governed by convenience in plotting dynamic characteristics using the method described in the preceding section. Hence, preference was given to the points of intersection of the static characteristics, $e_g = \text{constant}$, with the 10,000-volt E_p line. The explored dynamic characteristics for the UV-848 tube, corresponding to five selected operating points with I_o equal to 100,

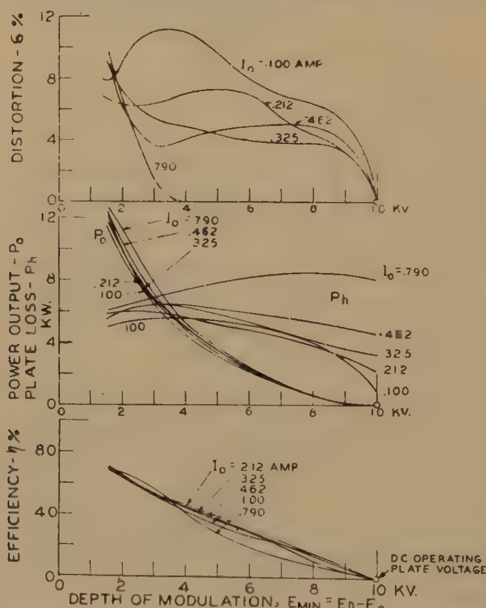


Fig. 14—Distortion as a function of modulation.
UV-863 tube
$E_p = 10.000$ volts
$R_L = 1680$ ohms
$\mu = 50$

250, 375, 525, and 810 milliamperes, are shown in Fig. 8. From these curves the transconductance characteristics given in Fig. 13 are plotted in the usual manner, and the calculation of distortion is carried out by the graphical method which has been outlined. Power output, plate loss, and efficiency are also computed for each operating line. The same procedure is followed for all four tubes.

The results of these calculations are graphically correlated in Figs. 14 to 17. All curves are plotted against the minimum plate voltage reached during modulation, given by the relation $E_{min} = E_p - E_o$ where E_p is the operating plate voltage and E_o is the amplitude of the alternating plate voltage corresponding to various depths of modulation.

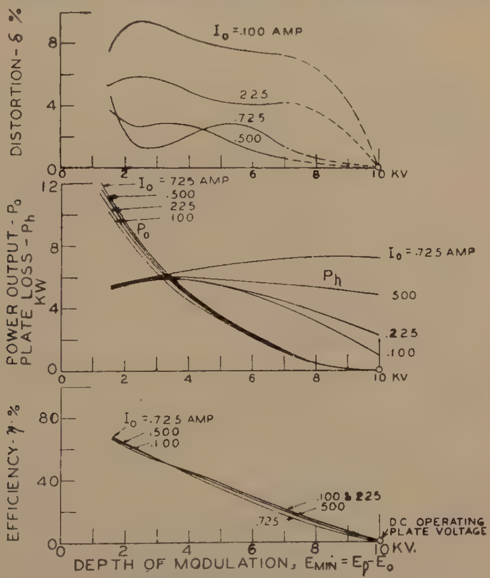


Fig. 15—Distortion as a function of modulation.
UV-207 tube
$E_p = 10{,}000$ volts
$R_L = 1680$ ohms
$\mu = 20$

Each figure contains four families of curves belonging to one particular tube. Individual curves in each family pertain to the chosen specific values of no-signal current; these values are indicated as parameters.

The upper group of curves in each figure records the variation of distortion along an explored dynamic characteristic as the depth of modulation is varied. Those in the middle give the output power, P_o, and the plate dissipation, P_h. The final graphs show the related values of efficiency, η per cent.

A thorough inspection of these curves and the comparison of analogous groups brings out the following conclusions:

(1) Distortion does not remain constant along any load line as the signal strength or depth of modulation changes. There is a general tendency toward decreased distortion as the signal strength is reduced. Yet, with high-μ tubes, such as the UV-863, and with medium values of no-signal current, the distortion can go up for lower modulations. This can be accounted for on the basis of the more complicated shape of the lower portion of the composite transconductance characteristics for such tubes. Incidentally, on high modulation this same cause is also responsible for the accentuated higher harmonics with the UV-863 tube. Tubes of this type may have distortion curves of a particularly irregular shape containing secondary maxima, minima, and inflection points.

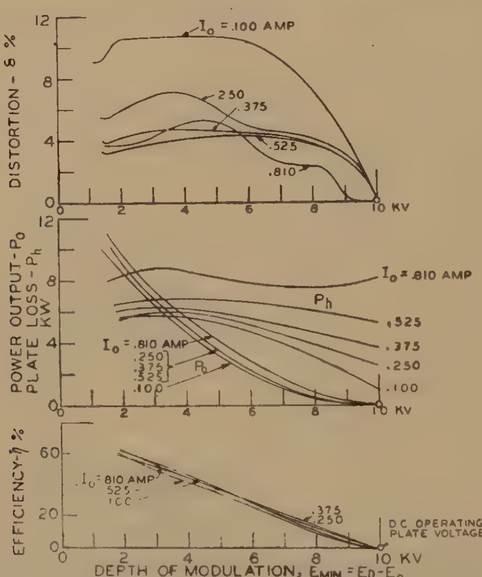


Fig. 16—Distortion as a function of modulation.
UV-848 tube
$E_p = 10{,}000$ volts
$R_L = 1680$ ohms
$\mu = 8$

(2) In general the distortion with all tubes decreases as the no-signal current is increased within the explored region. However, this rule does not hold strictly for all degrees of modulation. Under certain conditions governed by the exact shape of the transconductance characteristic there are local maxima or minima in the distortion as the no-signal operating point is shifted. This is conveniently illustrated in Fig. 18 in which the percentage of distortion, δ per cent, is plotted against the no-signal current, I_o, for all four tubes and for three different depths of modulation. The point with ten-kilowatt output for each tube is taken to correspond to 100 per cent modulation; on this

basis curves for 100, 50, and 33 per cent modulation are plotted. These curves indicate that the absolute minima of distortion in the various cases explored are generally located between no-signal currents of 500 milliamperes and one ampere, and even higher. Such operating conditions would depart essentially from those of an ideal class B amplifier and would border on class A operation with its inherently and

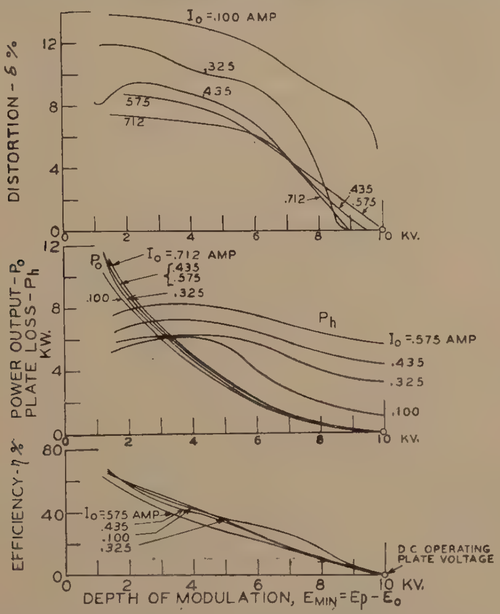


Fig. 17—Distortion as a function of modulation.
WL-419A tube
$E_p = 10{,}000$ volts
$R_L = 1680$ ohms
$\mu = 3.6$

unpleasantly high no-signal plate loss. However, taking into account the over-all results along any modulation path, one can establish an optimum value of I_o which is not inconsistent with class B operation, and which, if further increased, will produce only inappreciable improvement in operation. Thus, for example, from a detailed study of the curves of Fig. 18 one can arrive at a compromise between distortion and no-signal current as indicated in Table III.

TABLE III
DISTORTION WITH COMPROMISE BIAS

Type Type	μ	I_o	Distortion at:		
			100%	50%	33% Modulation
UV-863	50	0.3 amp	7.5%	4.0%	3.8%
UV-207	20	0.4	3.0	1.5	1.5
UV-848	8	0.4	4.0	4.2	3.2
WL-419A	3.6	0.45	9.0	6.0	3.0

(3) Referring again to the curves of Figs. 14 to 17 one can state that with all tubes, the power output, P_o, for the same minimum plate voltage increases slightly as the no-signal current is increased, due to the generally lower values of distortion at higher values of I_o.

(4) The plate loss is greater with higher no-signal currents. *This increase is much less pronounced on high modulation.* With decreasing signal amplitude the plate dissipation approaches the value of no-signal plate loss, which is, of course, proportional to I_o. As a rule, the maximum dissipation for all load lines corresponds to the condition of fifty per cent efficiency. For higher values of no-signal current the ab-

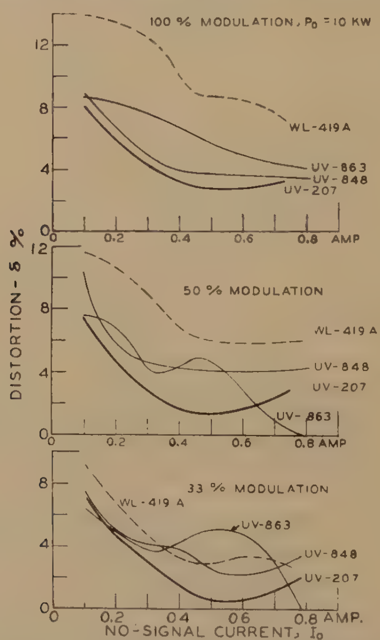


Fig. 18—Variation of distortion with no-signal current, I_0.
$E_p = 10{,}000$ volts
$R_L = 1680$ ohms

solute value of plate dissipation may exceed the six-kilowatt limit originally used for the plotting of the approximate dynamic load lines.

(5) For the same power output, the efficiency decreases but slightly as the operating point is shifted upwards. It is nearly the same for all tubes.

The Rôle of the Amplification Factor, μ, in Class B Audio Operation

From the foregoing analysis one can arrive at rather definite conclusions as to the effect of the amplification factor on distortion in class

B audio operation. However, a still better picture can be obtained by plotting the distortion against the amplification factor, as shown in Fig. 19. The three groups of curves refer, as before, to 100, 50, and 33 per cent modulation for a maximum power output of ten kilowatts. The value of no-signal current is used as a parameter. *All curves consistently show a definite minimum in the region between μ equal to fifteen and μ equal to twenty.* They also show that as the amplification factor drops below ten the distortion for similar operating conditions rises very rapidly.

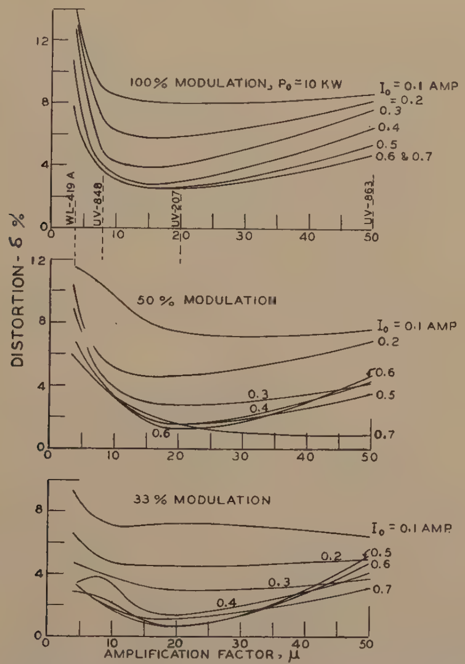


Fig. 19—Effect of amplification factor μ on distortion.
$E_p = 10{,}000$ volts $R_L = 1580$ ohms

It is interesting to point out that with the UV-863 tube having the highest μ, fifty, distortion can be substantially reduced by allowing a very high no-signal current of 790 milliamperes. With such a condition one is confronted with a permanent dissipation of eight kilowatts per tube. Yet, this represents an interesting case because the specified current corresponds *to zero bias operation.* With this condition, the defects introduced in class B operation by fluctuation of bias would cease to exist.

Plate Dissipation Limitations

A discussion of class B audio performance of water-cooled tubes in the light of permissible plate dissipation is quite essential for a com-

plete picture of this type of service. In Figs. 20 to 23 are shown the plate dissipations P_h, in kilowatts, which if exceeded, produce "hissing" and "boiling" of the cooling water at the anodes of each of the four tubes which have been the subjects of this investigation. These

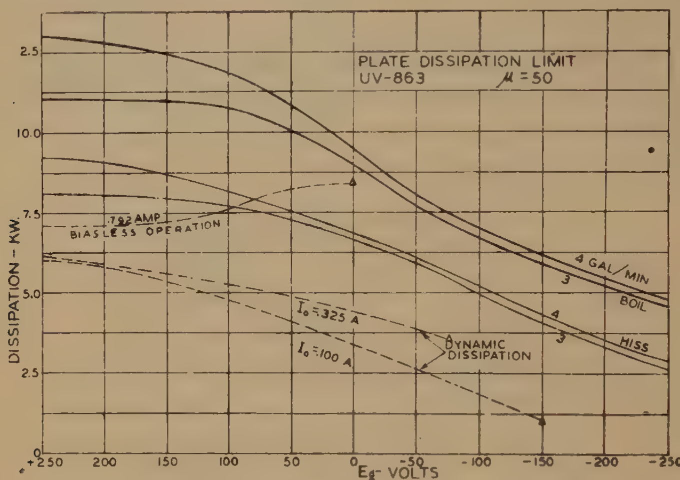


Fig. 20

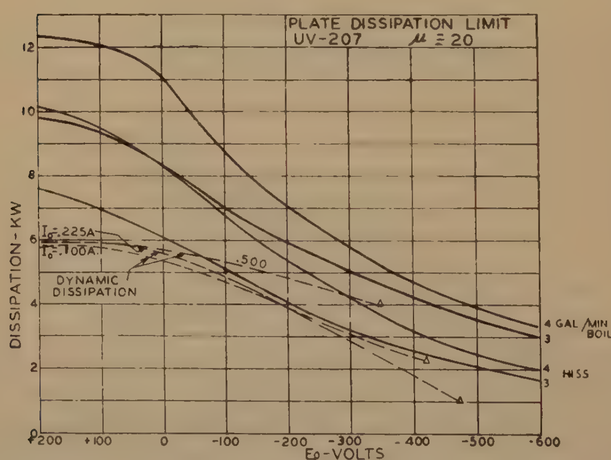


Fig. 21

values are plotted against grid voltage because this factor governs the width of the electronic beams arriving at the anode, and thus indicates the relative area of that portion of the total active anode surface in which heat is generated by the impinging electron beams. The beginning of both hissing and boiling was determined for each tube over the

indicated grid voltage ranges by means of a "stethoscope" made of a long piece of micarta tubing. One can either listen directly at the outer end of the stethoscope or attach to it a microphone and loud speaker system. Thus, at any chosen grid bias, the plate voltage is gradually increased until hissing and boiling become noticeable. The hissing, hav-

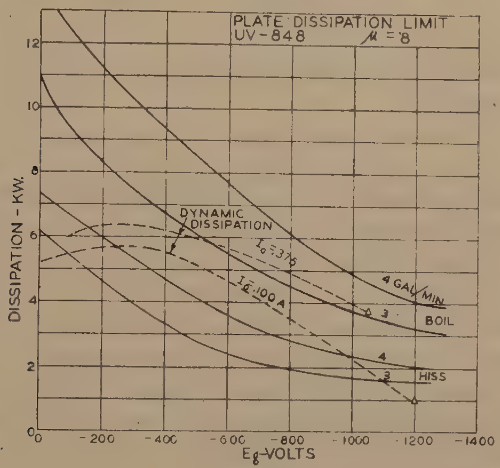


Fig. 22

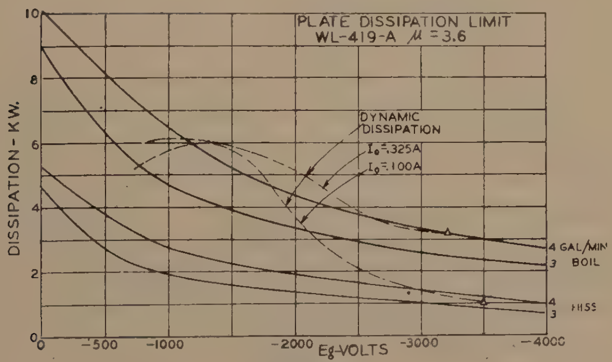


Fig. 23

ing a high audible pitch, indicates the beginning of the formation of minute steam bubbles, while the "boiling" gives a lower-pitched rattling noise accompanied by a slight shaking of the water jacket. The entire process of hissing and boiling has been observed visually by the use of a *glass water jacket.* The curves are plotted for three and four gallons per minute water flow.

As one might expect, low-μ tubes stand less dissipation in the condition of no-signal current. Thus, for grid voltages corresponding to the bias used with 10,000 volts on the plate, boiling starts

with UV-863	at P_{h_0}	6.5 kw
UV-207		4.5
UV-848		3.3
WL-419A		1.8

These values should be, and are, approximately inversely proportional to the square root of μ:

In the same charts of permissible plate dissipation operating curves are plotted showing the *actual* plate loss in operation along the six-kilowatt modulation path. This was calculated for various depths of modulation and is referred to the *average grid voltage* during the active half cycle of tube operation.

Evidently, for conservative operation, the curves of actual plate loss must not go over the curves of permissible plate dissipation, which, one might admit, is represented by the "hissing" rather than the "boiling" curves. Hissing and boiling bear no direct relation to the temperature of the outgoing water, being decidedly localized phenomena. From the data it is also clear that the lower the μ the more water is required for *efficient* cooling of the anode. One may feel that from this consideration alone tubes such as the WL-419A, having too low an amplification factor are practically excluded from use in class B audio service.

Grid Excitation Requirements

In order to compare the grid excitation requirements of the four tubes under investigation, the grid input power, P_g, has been calculated for each tube for two typical cases: a no-signal current of 0.100 ampere, and a no-signal current in the vicinity of 0.500 ampere. In all cases the plate voltage was assumed to be 10,000 volts, the power output, ten-kilowatt, and the load resistance 1680 ohms. The instantaneous values of grid currents for each tube, used in this calculation were obtained from complete grid-current charts plotted from oscillographic data. The results are assembled in Table IV.

TABLE IV

Tube Type	μ	No-Signal Current I_0	Grid Bias e_c	Direct Grid Current I_g	Grid Input P_g
UV-863	50	0.100 amp	− 150 volts	123 ma	92 watts
UV-207	20	0.100	− 475	47	55
UV-848	8	0.100	−1200	58	105
WL-419A	3.6	0.100	−3500	40	170
UV-863	50	0.462	− 50	92	71
UV-207	20	0.500	− 350	30	31
UV-848	8	0.525	−1000	35	54
WL-419A	3.6	0.435	−3100	19	70

The last column shows that with the output power chosen, and for both operating conditions, there is a distinct minimum in the required excitation power, located in the vicinity of μ equal to twenty. As the grid input power requirements play an important rôle in the design of the excitation stage, such a minimum is of considerable interest.

It may be pointed out that with power outputs of less than ten kilowatts, the necessary grid excitation drops more rapidly for low-μ tubes than for high-μ tubes. Thus, the excitation minimum for lower power outputs will shift in the direction of lower values of μ.

General Conclusions

The foregoing analysis deals with only a few typical cases of class B audio operation, identical for all four types of twenty-kilowatt tubes. Nevertheless, it is believed that, for other operating conditions, definite judgment in many particulars can be given concerning the comparative merits of the tubes which have been studied, using the results of this investigation as a point of perspective.

Thus, one can state that:

(1) For operation with the least distortion and with reasonably high power outputs under a wide variety of operating conditions, tubes having a medium value of amplification factor, μ, between fifteen and twenty-five, represent the best choice.

(2) High-μ tubes come next, because of their high plate dissipation ability, together with the ease and economy of bias supply. Thus they can be used at higher plate voltages, as well as at lower voltages in conditions approaching class A operation. *Biasless* operation of high-μ tubes is also quite feasible. However, one must be careful in employing these tubes for high outputs as the distortion rapidly rises to prohibitive magnitude.

(3) Low-μ tubes are the least suitable for the analyzed mode of operation. They *may* be utilized at lower plate voltages because under these conditions the plate loss limitation will not hamper the choice of no-signal current to such a degree as at higher voltages, and because the required grid bias is not impractically high. Power tubes with extremely low values of μ, say below six, can hardly be recommended for class B service in any circumstance unless one is concerned with *no grid current operation.*

One must keep in mind that this entire analysis has been carried out on the basis of tube charts mapped for an *individual* tube of each type. Therefore, for information regarding the *average* distortion values to be expected with any normal tube of a given type in class B service, one must explore charts of *average tube characteristics.*

By considering several operating plate voltages and assuming several load resistance values, one can prepare charts covering all feasible operating conditions for each type of tube. Such charts, giving a true and complete picture of class B tube performance, will allow the designer of new equipment to select the tube and operating conditions most fitted to his particular needs.

ADDENDUM

Suppose that a tube is to be used as a class B·audio amplifier. The operating point is given and the maximum positive voltage reached by the grid in operation is specified. The problem is: What is the slope of the load line for a maximum output?

As an example, let us refer to Fig. 1 belonging to the UV-863 tube. Let A be the operating point and AH one of the possible dynamic characteristics. With the grid swinging up to +700 volts (point H), power output, P_o, is proportional to the area of the triangle HAB; in fact, in the absence of distortion,

$$P_o = 1/4(\overline{OA} - \overline{OB}) \times \overline{HB} = 1/4(E_p - E_{\min})I_{\max}. \qquad (1)$$

For the point H, sliding along the line $e_g = \text{const} = +700$ volts, one can write

$$I_{\max} = f(E_{\min}). \qquad (2)$$

Then,

$$P_o = k(E_p - E_{\min}) \cdot f(E_{\min}). \qquad (3)$$

A maximum for P_o will be found in the usual way from

$$dP_o/dE_{\min} = 0.$$

Performing the indicated differentiation, one has

$$(E_p - E_{\min}) \cdot f'(E_{\min}) - f(E_{\min}) = 0.$$

Hence the condition for maximum output is

$$f'(E_{\min}) = f(E_{\min})/(E_p - E_{\min}) = I_{\max}/(E_p - E_{\min}). \qquad (4)$$

This can be written as

$$\tan \alpha = \tan \beta \qquad (5)$$

where α is the angle of the tangent to the curve for $e_g = \text{const}$, at the point H, and β is the angle between the load line and the X-axis. Thus (4) and (5) can be formulated in the following way: For a given operating point and a given crest grid voltage, the maximum output from a tube is obtained with a position of the load line on a plate-current /plate-voltage chart such that the angle formed by the load line and the

tangent to the curve along which the upper end of the load line slides, is bisected by the horizontal line passing through the vertex of the angle.

Designating the load resistance by R_L and the plate impedance at the point H by R_p, one may write for the condition of maximum output

$$R_L = R_p \tag{6}$$

because, according to the definition and construction,

$$R_p = 1/\tan\alpha \quad \text{and} \quad R_L = 1/\tan\beta.$$

The above derivation is not necessarily confined to the case of the curve $e_g = \text{const}$; it is valid for any curve plotted in the chart, such as, for example, the diode line $e_g = E_p$, as long as the end of the load line moves along the chosen curve and the power output can be considered as proportional to the area of the triangle HAB.

Proceedings of the Institute of Radio Engineers
Volume 23, Number 10 October, 1935

AN ELECTROMECHANICAL REPRESENTATION OF A PIEZO-ELECTRIC CRYSTAL USED AS A TRANSDUCER*

By
W. P. Mason
(Bell Telephone Laboratories, Inc., New York City)

Summary—*An electromechanical representation of a piezoelectric crystal which can represent the crystal when it is used to drive an external mechanical system is given in this paper. The elements of the system are evaluated in terms of the electrical, piezoelectric, and mechanical constants of the crystal. When the crystal is used as a purely electrical element, the representation reduces to the usual form. The equivalent network is applied in evaluating the effect of supersonic radiation on the decrement of a quartz crystal and in the design of a mechanical system driven by a piezoelectric crystal.*

Introduction

THE equivalent electrical representation of a piezoelectric crystal when used as an element in an electrical circuit has been discussed by several investigators,[1] who have arrived at the circuit shown by Fig. 1. Apparently, however, no simple circuit has been evolved when it is desirable to utilize the electromechanical coupling properties of the crystal to couple to an external system. Since such crystals are used[2] in loud speakers, microphones, and other apparatus,

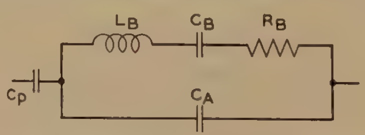


Fig. 1—Electrical representation of a piezoelectric crystal.

it is a matter of importance to obtain such a representation. This paper discusses such an equivalent circuit and relates the elements to the mechanical, electrical, and piezoelectric constants of the material. When used as a purely electrical element, this representation reduces to that of Fig. 1. Several examples of the use of this equivalent circuit are discussed.

* Decimal classification: 537.65. Original manuscript received by the Institute, April 18, 1934; revised manuscript received by the Institute, April 5, 1935; revised manuscript received by the Institute, July 30, 1935.

[1] W. G. Cady, *Phys. Rev.*, vol. 29, p. 1, (1922); Proc. I.R.E., vol. 10, pp. 83–115; April, (1922); K. S. Van Dyke, *Phys. Rev.*, Abstract, vol. 52, June, (1925); Proc. I.R.E., vol. 16, pp. 742–764; June, (1928); D. W. Dye, *Proc. Phys. Soc.*, vol. 38, (5), pp. 399–453; P. Vigoreaux, *Phil. Mag.*, pp. 1140–1153; December, (1928).

[2] A. M. Nicolson, *Proc. A.I.E.E.*, vol. 38, pp. 1315–1333, (1919); E. B. Sawyer, Proc. I.R.E., vol. 19, pp. 2020–2030; November, (1931); S. Ballantine, Proc. I.R.E., vol. 21, pp. 1399–1409; October, (1933).

An Electromechanical Representation of a Piezoelectric Crystal

Piezoelectric crystals are useful because they have the property of being deformed when voltage gradients are applied to them in particular directions. They may vibrate in many modes of motion but the two most used for driving external mechanical systems are longitudinal vibrations perpendicular to the applied electrical field and parallel to the applied field. Accordingly, the elements of the equivalent network are derived for these cases only. The network can, however, represent any type of motion driving a load, just as the network of Fig. 1 can represent the crystal for any type of motion. Methods for evaluating the constants of the network experimentally are discussed in the last section.

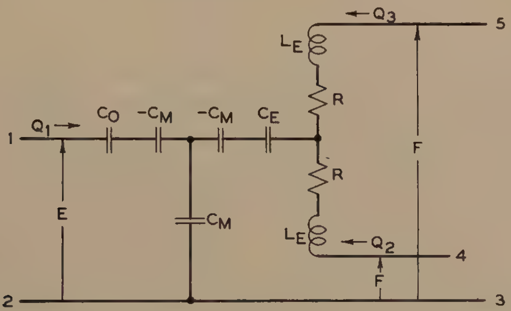


Fig. 2—Electromechanical representation of a symmetrical piezoelectric crystal.

Let us consider first the case of a crystal vibrating perpendicularly to the direction of the applied field. Two subdivisions of this case are usually of interest, the first when the crystal is supported at its center and drives two symmetrical loads, and the second when the crystal is supported on one end and drives a load on the other end. The symmetrical case is considered first.

By employing the well-known analogies between electrical and mechanical systems, it is possible to obtain a simple network, expressed in terms of electrical symbols, which portrays all of the properties of a piezoelectric crystal. In this representation, force is the analogue of voltage, mechanical displacement of electrical charge, and velocity of electrical current. When the electrodes are intimately associated with the crystal faces, the equivalent network of the crystal is shown by Fig. 2. To show this and to evaluate the elements comprising the network, let us consider several limiting cases. The voltage E is the voltage applied across the plates of the crystal, and the force F is the force applied to each end of the symmetrical crystal. The charge

Q_1 is the electrical charge flowing through the wires connected to the crystal, and Q_2 and Q_3 are the mechanical displacements of the ends of the crystal. On account of the symmetry Q_2 equals Q_3. Suppose now that a voltage E is applied to a clamped crystal. Then since $Q_2 = Q_3 = 0$, the charge collected on the condenser C_0 is equal to

$$Q_1 = C_0 E \tag{1}$$

since the remainder of the circuit carrying charge consists of a positive and a negative C_M which cancel each other. Hence C_0 is the electrostatic capacitance of the crystal when it is clamped. If K is the dielectric constant of the crystal clamped then C_0 in c.g.s. electrostatic units is

$$C_0 = \frac{K l_m l_0}{4\pi l_e} \tag{2}$$

where l_e is the dimension of the crystal in centimeters perpendicular to the surfaces of the electrodes, l_m the length of the crystal in the direction of vibration, and l_0 the length of the third axis.

Since the same charge Q_1 flows through the shunt condenser C_M, the force F required to keep the crystal from moving is obviously

$$F = Q_1/C_M \text{ or } C_M = Q_1/F \text{ (unit is charge per dyne)}. \tag{3}$$

The mutual capacitance-compliance C_M might be defined as the ratio of the charge applied to the crystal to the force required to keep it from moving. Since, however, no piezoelectric constant is measured in this way, another derivation of C_M is given later, involving the usual piezoelectric constant d, which is measured by observing the amount of expansion of the crystal for an applied potential gradient.

Now let us apply a static force, symmetrically, on each end of the crystal, and leave the electrical plates open-circuited. Then the crystal acts as though the plates were not there (i.e., the plates could be removed without affecting the displacement) and hence the displacement of the crystal can be calculated from the elastic constant of the crystal. For a given force F applied to the crystal, the total displacement is

$$Q_2 + Q_3 = \frac{F s l_m}{l_e l_0} \tag{4}$$

where s is the modulus of compliance of the crystal (the inverse of Young's modulus) along the axis of vibration. But $s l_m / l_e l_0$ is the mechanical compliance of the crystal, in c.g.s. units, which has been designated by C_E, or

$$C_E = \frac{sl_m}{l_e l_0} \text{ (displacement per dyne)}. \tag{5}$$

In a similar way we find that

$$C_M = (Q_2 + Q_3)/E \text{ (displacement per unit c.g.s. electrostatic potential)} \tag{6}$$

that is, C_M is the ratio of the displacement to the open-circuit electrical potential. Hence, the network of Fig. 2 displays the static electrical and mechanical properties of the crystal.

Suppose now that an alternating force is applied to the crystal. Then since the crystal has inertia it will exert a mass reaction to an applied force, which is represented by the mechanical inductance L_E. At low frequencies this inductance in mechanical impedance units will be half the mass of the crystal. However, near the resonance of the crystal this equivalent inductance will be less due to the fact that the crystal does not move as a whole, but has a velocity which is proportional to the sine of the distance of a given point from the center of the crystal. Hence the total kinetic energy is less and the equivalent value of L_E smaller.[3] Mechanical resonance in the network will occur when L_E is in resonance with half the compliance C_E since identical forces F are acting on the two ends of the symmetrical crystal. Since the mechanical frequency of the resonance of such a rod is

$$f_A = \frac{1}{2l_m\sqrt{\rho s}} \tag{7}$$

where ρ is the density of the crystal, it follows that

$$L_E = \frac{2l_e l_m l_0 \rho}{\pi^2} \text{ (dynes per unit acceleration)}. \tag{8}$$

The resistances R shown include the dissipation due to internal friction, supersonic radiation from the ends of the crystal, friction at the point of support, and all other sources of dissipation.

Hence all of the elements of the network are determined in terms of the mechanical and electrical constants, except the mutual capacitance-compliance C_M. In order to evaluate C_M use is made of the piezoelectric equation

$$e = dV \tag{9}$$

[3] Strictly speaking, the value of C_E is also a function of frequency, but at the first resonance it can be shown that it differs from its static value by the factor $8/[\pi^2 - k^2(\pi^2 - 8)]$. For a highly coupled crystal this factor does not differ much from unity, and hence in the interest of simplicity the variation of C_E is neglected.

where e is the strain (per unit length) produced in the crystal by the applied potential gradient V and d (the constant of proportionality) is the piezoelectric coefficient. Solving for the charge and the displacement at zero frequency for the network of Fig. 2, we find by employing Kirchoff's laws

$$Q_1 = \frac{EC_0 + Fk\sqrt{C_0C_E}}{1 - k^2}; \qquad (Q_2 + Q_3) = \frac{Ek\sqrt{C_0C_E} + FC_E}{1 - k^2} \tag{10}$$

where k, the coefficient of coupling, is defined by

$$k^2 = \frac{C_0C_E}{C_M{}^2}. \tag{11}$$

In order to evaluate k, the coefficient of coupling between the electrical and mechanical systems, in terms of d, let F (the force) vanish since d is the strain per unit length per unit potential gradient for a free crystal. Then,

$$(Q_2 + Q_3) = \frac{Ek\sqrt{C_0C_E}}{1 - k^2}. \tag{12}$$

Comparing this with (9), noting that $(Q_2+Q_3)=el_m$, and $V=E/l_e$, we have on the insertion of the values for C_0 and C_E from (2) and (5)

$$d = \frac{k\sqrt{\dfrac{Ks}{4\pi}}}{1 - k^2}. \tag{13}$$

Solving for k we find

$$k = \frac{1}{2d}\sqrt{\frac{Ks}{4\pi}}\left[-1 + \sqrt{1 + \frac{16\pi d^2}{Ks}}\right]. \tag{14}$$

If $16\pi d^2/Ks$ is a small quantity, as it is in quartz, for example, this reduces approximately to

$$k \doteq d\sqrt{\frac{4\pi}{Ks}}. \tag{15}$$

However, for any crystal with a large coupling (such as Rochelle salt which may have a coupling of 75 per cent) (15) is not very accurate.

When the crystal is used as an element in an electrical network, and allowed to vibrate freely, the force F of Fig. 2 can be set equal to zero, and the network short circuited. Solving for the impedance on the electrical side we find

$$Z_C = \frac{-j(1-k^2)}{2\pi f C_0}\left[\frac{1-f^2/f_1^2+j/q(1-k^2)}{1-f^2/f_2^2+j/q}\right] \tag{16}$$

where $f_2^2=f_A^2$; $f_1^2=f_A^2\,(1-k^2)$, (f_A being the natural mechanical resonance frequency of the crystal) and q is the ratio of the reactance of the condenser C_E to the resistance $R/2$ or

$$q = 2/R\omega C_E. \tag{17}$$

It is easily shown that the impedance Z_C is also the impedance of the network of Fig. 1 (with C_p set equal to infinity since the plates are in intimate contact with the crystal) if

$$C_A=C_0;\quad C_B=C_0k^2/(1-k^2);\quad L_B=1/4\pi^2 f_A^2 k^2 C_0;\quad R_B=1/\omega C_0 k^2 q. \tag{18}$$

Hence, the representation in Fig. 2 reduces to the well-known Fig. 1, when the crystal is free to vibrate. It will be noted from (18) that at a low frequency, a capacitance C_B is added to the purely electrostatic capacitance C_A, the added capacitance being due to the storing up of mechanical energy by means of the piezoelectric effect. In case the electrodes are not in intimate contact with the crystal faces an added capacitance C_p is placed in series with the condenser C_0. This capacitance is due to the air gap and may be calculated by the formula

$$C_p = \frac{l_m l_0}{4\pi(t_1+t_2)} \tag{19}$$

where t_1 and t_2 are the thicknesses of the air gaps separating the electrodes and the crystal. This network then reduces identically to Fig. 1.

A network representing the second case, when one end is supported and the other end used to drive a load, is shown on Fig. 3. The method of deriving the constants is the same as before and the only difference comes in the determination of the mechanical inductance L_E which is different due to the difference in the resonance frequency. For a bar clamped on one end and driven by a force F, the resonant frequency is given by

$$f_A = \frac{1}{4l_m\sqrt{s\rho}} \tag{20}$$

Since the mechanical inductance L_E resonates with the compliance C_E we find

$$L_E = \frac{4\rho l_m l_e l_0}{\pi^2} \tag{21}$$

which is twice the inductance found before and corresponds to the fact that twice the mass is moved from the clamping position.

When the direction of motion is parallel to the direction of the applied field, the same networks hold but the elements have different lengths entering into their determination. The direction of the applied field and of the motion is designated by l_e. The other two axes are still designated by l_m and l_0. The resulting constants are

$$C_0 = \frac{K l_m l_0}{4\pi l_e}; \qquad C_E = \frac{s l_e}{l_m l_0}; \qquad C_M \doteq \frac{Ks}{4\pi d};$$

$$L_{E_1} = \frac{2\rho l_e l_m l_0}{\pi^2}; \qquad L_{E_2} = \frac{4\rho l_e l_m l_0}{\pi^2} \tag{22}$$

where L_{E1} is the mechanical inductance for the symmetrical case (Fig. 2) and L_{E_2} for the dissymmetrical case (Fig. 3).

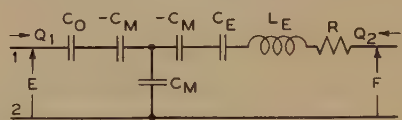


Fig. 3—Electromechanical representation of a piezoelectric crystal clamped on one end.

Application of Equivalent Circuit to Electromechanical Systems

A simple example of the use of Fig. 2 in determining the effect of a mechanical load on the impedance of a crystal is the problem of finding out how much the energy radiated by a crystal to the surrounding air affects the decrement of a quartz crystal vibrating longitudinally. When a crystal vibrates longitudinally, energy is radiated to the surrounding medium by the motion of the ends of the crystal. If the dimensions of the ends of the crystal are comparable to a wavelength or greater—which they will ordinarily be for a crystal vibrating at a high frequency—it is well known[4] that the radiating surface will experience a resistance to motion equal to

$$R_R = \rho_A b \text{ (mechanical ohms)} \tag{23}$$

per square centimeter, where ρ_A is the density of the medium and b the velocity of propagation. For air R_R is about 41 ohms per square centimeter. Hence, the equivalent circuit for a crystal vibrating in air is Fig. 2 terminated at the terminals 3–4 and 3–5 by mechanical resistances

$$R_R = 41 l_t l_0. \tag{24}$$

[4] *See* I. B. Crandall, "Theory of Vibrating Systems and Sound," Chapter 4, B. Van Nostrand Co.

The remaining resistances R of Fig. 2 will then include all other sources of dissipation, such as clamping resistance, internal frictional resistance, etc.

If all these other sources of dissipation were eliminated the radiation resistance would produce a limiting value for the decrement of the crystal which may be calculated as follows: From (17) the value of q for the mechanical circuit is

$$q = \frac{2}{2\pi f R_R C_E} = \frac{2}{2\pi \times 41 s l_m f} \tag{25}$$

on inserting the value of R_R and C_E. At the mechanical resonance of the crystal $f = 1/2\, l_m\sqrt{\rho s}$, hence

$$q = \frac{2\sqrt{\rho/s}}{41\pi}. \tag{26}$$

Since $\rho = 2.65$ and $s = 1.27 \times 10^{-12}$ for a perpendicularly cut quartz crystal

$$q = 2.24 \times 10^4. \tag{27}$$

The decrement of a crystal in terms of the circuit of Fig. 1 is

$$\delta = \frac{R_B}{2 f_A L_B}. \tag{28}$$

Inserting the values of (18) we find[5]

$$\delta = \frac{\pi}{q} = 1.4 \times 10^{-4}. \tag{29}$$

To get the much lower decrements reported by Van Dyke,[6] it is necessary to put the crystal in an evacuated chamber, where the energy radiated into the surrounding medium is zero. It should be noted that this calculation does not apply to any other mode of motion than the longitudinal mode because the radiating efficiencies of the different modes of motion are not the same. For example all of the motion of a crystal vibrating in shear is tangential to the face of the crystal and hence such a crystal should not radiate appreciably. However, due to

[5] If account is taken of the variation of C_E with frequency discussed in footnote 3, the limiting value of q is 2.77×10^4 and the limiting decrement is 1.13×10^{-4}.

[6] *Phys. Rev.*, November 15, (1934), Abstract (8) and Proc. I.R.E., vol. 23, pp. 386–393; April, (1935). In the last-mentioned reference Van Dyke gives a limiting value of decrement of 1.26×10^{-4} for a bar vibrating freely in air. Since the residual losses were about five per cent of the radiation losses, this value agrees well with the value given in footnote 5.

the fact that shear motion is coupled to flexural motion, there will be some displacement normal to the surfaces of the crystal and hence some radiation. For example, the decrement of the "doughnut" crystal such as used in the Bell System frequency standard is in the order of 3×10^{-5} when mounted freely in air, the principal vibration being a shear coupled to a flexure.

The equivalent circuits of Figs. 2 and 3 may also be used as the basis of design for mechanical systems, such as loud speakers, which are driven by crystal elements, and taken together with network theory can be used to obtain the proper dimensions of a crystal to perform a definite function. Since the electrical impedance of the element will not ordinarily be equal to the mechanical impedance of the mechanical system that the crystal is driving, a modification of the circuits of Figs. 2 and 3 employing perfect transformers is very useful. As shown by E. L. Norton,[7] a perfect transformer shunted by an impedance Z_s is equivalent to a T network having the values shown on Fig. 4. To apply this identity to Fig. 3, let the impedance transformation ratio ϕ^2 equal C_0/C_E. Then since $C_M=\sqrt{C_0C_E/k}$ it follows that $Z_s=-(jk/\omega C_0)$. To make up the impedance Z_A it is necessary to break the capacitance C_0 into two capacitances in series, $C_0/(1-k)$ and C_0/k and for Z_B to break the compliance C_E into the compliances $C_E/(1-k)$

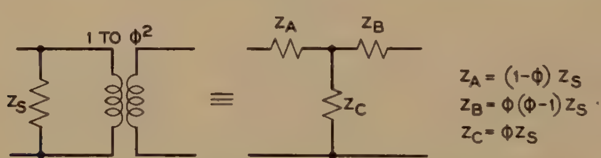


Fig. 4—Equivalent network for an ideal transformer shunted by an impedance

and C_E/k in series. Then C_0/k and $-C_M$ make up the impedance Z_A, while C_E/k and $-C_M$ make up the impedance Z_B, with the result that the network of Fig. 3 can be transformed into that of Fig. 5. All elements on one side of the transformer are electrical, while those on the other side are mechanical. Hence the transformer transforms from electrical impedance to mechanical impedance. The shunt capacitance C_0/k can be placed on the other side of the transformer if desired in which case it will have a value C_E/k. A similar transformation can be applied to the network of Fig. 2.

As an example of the use of such a representation, a recent paper[2] by Ballantine describes a high-frequency loud speaker system driven

[7] U.S. Patent 1,681,554, August 21, 1928; also T. E. Shea, "Transmission Networks and Wave Filters," page 326, B. Van Nostrand Co.

by a Rochelle salt diaphragm. It is found that a coil in series with the crystal increases the band width of the response. The reason for this will be obvious if the circuit of Fig. 5 is used to represent the crystal.

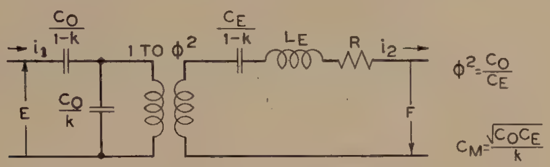


Fig. 5—Electromechanical representation of a piezoelectric crystal, employing an ideal transformer.

The other elements of the circuit, introduced by the loud speaker system, are a shunt compliance C_C, representing the compliance of the throat chamber, and a terminating resistance R_H due to the radiation resistance of the horn, whose values in mechanical impedance units are

$$C_C = 7 \times 10^{-7}\, t/S_d; \qquad R_H = 41 S_d^2/S_t \qquad (30)$$

where t is the clearance between the diaphragm and the base of the horn, S_d the effective area of the diaphragm, and S_t the area of the horn throat. Fig. 6 shows the complete network for the crystal and loud speaker. The network has the form of a well-known band-pass filter

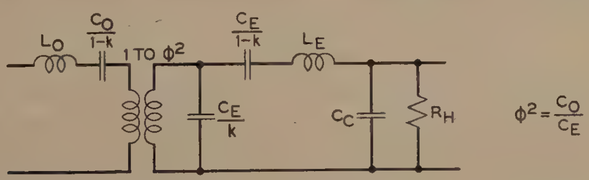


Fig. 6—Electromechanical representation of a loud speaker.

and the added electrical inductance L_0 completes the termination on the electrical side. Without this inductance the network represents a dissymmetrical band-pass filter having a much narrower pass band. Since the problem of a crystal driving a loud speaker is typical of a number of problems occurring when crystals drive mechanical systems, it has been suggested by the board of editors that this example be worked out in detail.

The elements of the network representing the crystal diaphragm are not easily derived from the fundamental constants due to the complexity of the mode of motion employed and due to the fact that the load is applied in a distributed manner over the complete surface of the crystal. The elements can, however, be measured experimentally by electrical or mechanical measurements. If the diaphragm is placed in the

open air (or better still in a vacuum to eliminate the effects of radiation) and its electrical impedance measured over a range of frequencies, the electrical capacitance C_0 and the coefficient of coupling k can be evaluated. Since the crystal is free to move, the network representing its impedance is that shown on Fig. 1 with C_p equal to infinity since the electrical plates are attached to the surface of the crystal. At low frequencies, the capacitance measured is C_A+C_B which from (18) is

$$C_A + C_B = C_0/(1 - k^2). \tag{31}$$

Next the resonant and antiresonant frequencies f_1 and f_2 can be measured by measuring the frequencies of maximum and minimum current. As shown by (16)

$$f_1^2/f_2^2 = 1 - k^2 \quad \text{or} \quad k = \sqrt{1 - f_1^2/f_2^2}. \tag{32}$$

Furthermore the frequency of mechanical resonance of the plated crystal f_A is equal to the antiresonant frequency of the crystal f_2, so that

$$L_E = 1/4\pi^2 f_A^2 C_E. \tag{33}$$

Hence, all of the elements are determined except the mechanical compliance of the crystal C_E, which determines the ratio of transformation.

To measure C_E either a mechanical measurement or a mechanical modification of the system is required. A simple means of measuring C_E is to close off the diaphragm chamber, which is equivalent to terminating the network of Fig. 5 in the known compliance C_C of (30). The mechanical system will then have a higher frequency of resonance f_A' which can be measured electrically by measuring the antiresonant frequency f_2'. The compliance C_E can be calculated from the formula

$$C_E = C_C\left[\left(\frac{f_A'}{f_A}\right)^2 - 1\right]. \tag{34}$$

Hence all of the elements of the network can be determined experimentally for a given diaphragm thickness. Both C_0 and C_E are inversely proportional to the diaphragm thickness.

When the electrical inductance L_0 is placed in series with the crystal, the network of Fig. 6 has the form of a well-known[8] band-pass filter. In terms of R_H, which is a mid-shunt impedance, the mechanical elements have the following values:

[8] *See* T. E. Shea, "Transmission Networks and Wave Filters," page 283.

$$L_E=\frac{f_{c_2}R_H}{\pi(f_{c_2}^2-f_{c_1}^2)};\ \frac{C_E}{1-k}=\frac{f_{c_2}^2-f_{c_1}^2}{4\pi f_{c_1}^2 f_{c_2}R_H};\ C_C=\frac{1}{4\pi f_{c_2}R_H};\ \frac{C_E}{k}=\frac{1}{\pi f_{c_2}R_H}\tag{35}$$

where f_{c_1} and f_{c_2} are the lower and upper cutoff frequencies of the filter. Their ratio is determined when the coupling factor k is known, for upon taking the ratio of $C_E/(1-k)$ to C_E/k, we find

$$\frac{f_{c_2}}{f_{c_1}}=\sqrt{\frac{1+3k}{1-k}}.\tag{36}$$

Next taking the product of L_E and $C_E/(1-k)$, we find

$$f_{c_1}=f_A\sqrt{1-k}\ \text{ so that }\ f_{c_2}=f_A\sqrt{1+3k}\tag{37}$$

and hence the upper and lower cutoff frequencies of the band which will be radiated by the loud speaker are determined by the mechanical resonance of the crystal f_A and the coupling coefficient k. The dimensions of the crystal should be so adjusted that

$$C_E = 4k/4\pi f_A R_H\sqrt{1+3k}\tag{38}$$

and the thickness of the air chamber so that

$$C_C = 1/2\pi f_A R_H\sqrt{1+3k}\tag{39}$$

which determine all of the mechanical constants of the system.

Turning now to the electrical portion of the circuit of Fig. 6, it would be desirable to terminate the filter at its mid-series impedance. This, however, cannot be done because when the transformation ratio C_E/C_0 is fixed, C_0 has the right value for a full series termination. A fairly good result is obtained when the value of the inductance L_0 resonates the capacitance C_0 at f_A, i.e.,

$$L_0 = 1/4\pi^2 f_A^2 C_0\tag{40}$$

and the electrical impedance that the loud speaker works out of is taken as the mid-series impedance of the filter which is equal to

$$R_I=\frac{\sqrt{1+3k}-\sqrt{1-k}}{4\pi f_A C_0}.\tag{41}$$

This corresponds to a filter with a full series termination working out of its mid-series impedance. A better result will be obtained by completing the electrical termination of the filter with a shunt antiresonant circuit in which the coil and condenser have the values

$$L_s=\frac{R_I(f_{c_2}-f_{c_1})}{2\pi f_{c_1}f_{c_2}};\qquad C_s=\frac{1}{2\pi(f_{c_2}-f_{c_1})R_I}.\tag{42}$$

Proceedings of the Institute of Radio Engineers
Volume 23, Number 10 *October, 1935*

DISCUSSION ON "CONTROL OF RADIATING PROPERTIES OF ANTENNAS"*

C. A. Nickle, R. B. Dome, and W. W. Brown

G. H. Brown:[1] The authors have shown an interesting method of shifting the current distribution along an antenna. They have enumerated the advantages of such a system and show that it is very desirable to shift the current distribution so that a nodal point occurs at the base of the antenna.

I believe that the authors must have used very small ground systems, not over one twentieth of one wavelength in radius. When adequate ground systems are used, the reduction in earth loss due to use of the tuned capacity area is less important. It is, however, still possible to get large gains in field strength at the horizon due to redistribution of the energy in the vertical plane.

Let us first consider Fig. 4 of the paper under discussion. Curve (1) shows the field strength vs. antenna height for a vertical wire antenna when the power input is held constant. This curve shows that the field intensity increases very rapidly with antenna height. We see that an antenna one-half wave tall gives a signal about fifty-seven per cent greater than does an antenna one-quarter wave tall. This is surprising in view of the fact that theoretical considerations set this figure at twenty-one per cent when there are no power expenditures other than radiation. I have made calculations which show that for average earth in the broadcast band the ground system consisting of sixty or more radials must be under twenty feet in radius to yield a difference of fifty-seven per cent in the field intensities. P. P. Eckersley, T. L. Eckersley, and H. L. Kirke[2] made similar measurements on quarter- and half-wave antennas, using a ground system consisting of forty radial wires each 250 feet long, buried six inches deep in dry soil over a chalk base. They report an increase of twenty-six per cent in changing from the quarter- to the half-wave antenna. Their measurements of resistance show that the efficiency of the quarter-wave antenna was 82.3 per cent while the efficiency of the half-wave antenna was 88.2 per cent.

One of the examples discussed by the authors is an eighth-wave long antenna. They predict a large increase in field intensity when sufficient loading is placed at the top of the antenna to cause a current node to occur at the base of the antenna. It is interesting to consider the magnitude of the earth currents in the neighborhood of the antenna. Let us call the simple eighth-wave wire antenna A, and the antenna of the same height but with sufficient loading at the top to cause a current node at the bottom will be designated as antenna B. Let I_x be the total earth current flowing radially inward toward the base of the antenna across a circle of radius, x, whose center coincides with the base of the antenna. The following table shows the values of this earth current as a function of the distance, x, when the radiated power in each case is considered to be 1000 watts.

x (*wavelengths*)	I_x (*amps*) Antenna A	I_x (*amps*) Antenna B
0.0	12.0	0.0
0.025	10.45	2.15
0.05	9.15	3.62
0.1	7.5	5.0
0.2	6.05	5.6
0.3	5.42	5.5
0.5	5.1	5.1

* Proc. I.R.E., vol. 22, no. 12, pp. 1362–1373; December, (1934).
[1] RCA Victor Company, Inc., Camden, New Jersey.
[2] P. P. Eckersley, T. L. Eckersley, and H. L. Kirke, "The design of transmitting aerials for broadcasting stations," *Jour.* I.E.E. (London), vol. 67, p. 507, (1929).

Thus we see that at distances greater than one-tenth wavelength from the base of the antenna the earth current is essentially the same for both antennas. We may thus conclude that if a goodly number of radial wires are extended to at least one tenth of a wavelength from the base of the antenna there would be little or no advantage in causing a node of current to occur at the base of the antenna.

The authors also show by means of curve (3) Fig. 4, that it is possible to adjust the top loading on a quarter-wave antenna so that the field strength is increased by forty-five per cent. Their theoretical considerations show that placing the current node at the bottom of the antenna causes a six per cent increase due to decreased radiation resistance. They then conclude that the forty-five per cent increase obtained experimentally is made up of this six per cent gain plus a remaining thirty-nine per cent due to decreased ground connection loss. In describing their experimental results, the authors expressly state that the sphere and inductance combination was "carefully tuned for maximum field strength." It does not follow that this point coincides with the condition where the current node occurs at the base of the antenna. In fact, if one considers the case where the current node occurs at a point 0.111 wavelength from the ground, it is found that the field strength is increased thirty-eight per cent due to the changed radiation characteristic. We might then conclude that the ground losses account for the remaining seven per cent increase. It is worth mentioning that the high angle radiation is reduced a great deal more in this condition than it is for either condition shown in Fig. 2.

The curves of Fig. 5 again show the influence of a very small ground system. Curve (1) of this figure is not at all in agreement with the results obtained on actual tower installations where an adequate ground system is used.

When large ground systems are used, one is forced to conclude that the reduction in earth loss through the use of the tuned top becomes small. However, the tuned arrangement is still important because of its effects on the distribution of radiation. A more detailed account of this effect has been published elsewhere.[3]

It is important to note that the radiation resistance of the antennas less than one-third wavelength long is reduced to very small values when the top loading is adjusted to give the maximum suppression of sky wave. Then only a few ohms of resistance in coupling and loading coils will seriously reduce the efficiency of the antenna.

C. A. Nickle, R. B. Dome, and W. W. Brown:[4] The ground system used in the tests consisted of eight radial wires, each 0.08 inch in diameter and 135 feet long, spaced forty-five degrees apart around the base of the antenna. These were laid upon the ground with each end soldered to a copper pipe driven into the earth. It is agreed that the number of wires could have been increased to reduce further the ground resistance.

[3] H. E. Gihring and G. H. Brown, "A brief survey of the characteristics of broadcast antennas," *Broadcast News*, p. 8, December, (1934).
[4] General Electric Company, Schenectady, New York.

BOOKLETS, CATALOGS, AND PAMPHLETS RECEIVED

Copies of the publications listed on this page may be obtained without charge by addressing the publishers.

National Union Laboratories of 365 Ogden St., Newark, N. J., are issuing a series of technical pamphlets on vacuum tubes. The first nine of the series are as follows: 6A8G, Self-Excited Electron-Coupled Converter; 6C5G, Medium-Mu Voltage Amplifier; 6D5G, Power Amplifier; 6F5G, High-Mu Voltage Amplifier; 6F6G, Power Amplifier; 6H6G, Twin Diode; 6J7G, Sharp Cut-Off Detector and Amplifier; 6K7G, Remote Cut-Off Radio and Intermediate-Frequency Amplifier; and 5Y3, Full Wave Rectifier.

Rheostats and resistance units are described in Catalog 14 issued by the Ohmite Manufacturing Company, 636 N. Albany Ave., Chicago, Ill.

Technical data and suggested uses of its laminated phenolic material under the name of Dilecto are given in a booklet issued by the Continental-Diamond Fibre Company, Newark, Del.

Model 73A, Crystal Lapel Microphone, is described in a leaflet issued by Shure Brothers Company of 215 W. Huron St., Chicago, Ill. Model 70S, Communications Type Crystal Microphone, is described in another leaflet.

Trimmer condensers are described in Bulletin No. 38 issued by DeJur-Amsco Corporation of 95 Morton St., New York City.

RCA Manufacturing Company of Harrison, N. J. has issued Application Note No. 50 on the operation of the 6L7 as a mixer tube. Technical booklets have been issued on the 803 Radio-Frequency Power Amplifier Pentode and the 838 Class B Modulator, Radio-Frequency Power Amplifier, Oscillator. A single bulletin on "New All-Metal Radio Tubes" gives data on the 5Z4, Full Wave High Vacuum Rectifier; 6A8, Pentagrid Converter; 6C5, Detector Amplifier Triode; 6F5, High-Mu Triode; 6F6, Power Amplifier Pentode; 6G6, Twin Diode; 6J7, Triple-Grid Detector Amplifier; 6K7, Triple-Grid Super-Control Amplifier; and 6L7, Pentagrid Mixer Amplifier.

CONTRIBUTORS TO THIS ISSUE

Burnside, D. G.: Born January 1, 1891, at Point Pleasant, West Virginia. Engineering department, Atwater Kent Company, 1926–31, research and development laboratory, RCA Radiotron Company, Inc., 1931 to date. Associate member, Institute of Radio Engineers, 1912.

Herold, E. W.: Born October 15, 1907, at New York City. Received B.Sc. degree, University of Virginia, 1930. Western Electric Company, 1924–1925; Bell Telephone Laboratories, Inc., 1925–26; engineering department, E. T. Cunningham, Inc., 1926–1929; research and development laboratory, RCA Radiotron Company, Inc., 1930 to date. Associate member, Institute of Radio Engineers, 1930.

Jansky, Karl G.: Born October 22, 1905, at Norman, Oklahoma. Received A.B. degree, University of Wisconsin, 1927. Bell Telephone Laboratories, Inc., 1928 to date. Associate member. Institute of Radio Engineers, 1928; Member, 1934.

Kirby, S. S.: See PROCEEDINGS for July, 1935.

Kozanowski, Henry N.: See PROCEEDINGS for July, 1935.

Kramar, Ernest A. Ludwig: Technical High School, Prague, 1919–1925; Technical High School, Dresden, laboratories of Professor Barkhausen, 1925–1927; received Dr. Ing. degree, 1927. Radio laboratories, C. Lorenz Aktiengesellschaft, Berlin-Tempelhof, 1927 to date. Member, W. G. L. Nonmember, Institute of Radio Engineers.

Lester, G. H.: Born January 6, 1904, at Auburn, New York. Graduated Bliss Electrical School, 1926. Westinghouse Electric and Manufacturing Company, 1926–1930; National Bureau of Standards, 1930 to date. Nonmember, Institute of Radio Engineers.

Mason, W. P.: Born September 28, 1900, at Colorado Springs, Colorado. Received B.S. degree, University of Kansas, 1921; M.A. degree, Columbia University, 1924; Ph.D. degree, 1928. Engineering department, Western Electric Company, 1921–1925; Bell Telephone Laboratories, Inc., 1925 to date. Member, Sigma Xi; American Physical Society; Fellow, Acoustical Society of America. Nonmember, Institute of Radio Engineers.

Morgan, Howard K.: Born February 22, 1906, at New York City. Received B.S. degree, University of California, 1929. Radio test course, General Electric Company, 1929–1930; special receiver engineering, RCA Victor Company, 1930–1931; test apparatus development engineering, International Resistor Company, 1931; receiver development and production engineering, Pilot Manufacturing Company, 1931; receiver development and aircraft experimental radio engineering, Transcontinental and Western Air, Inc., 1932 to date. Associate member, Institute of Radio Engineers, 1927.

Mouromtseff, Ilia Emmanuel: See PROCEEDINGS for July, 1935.

Norton, K. A.: Born February 27, 1907, at Rockwell City, Iowa. Received B. S. degree, University of Chicago, 1928. Insp-ction development laboratory. Western Electric Company, 1929. Columbia University, 1930-1931. Rodio Section, National Bureau of Standards, July, 1929, to September, 1930, and June, 1931, to December, 1934. Federal Communications Commission, December, 1934 to date. Associate member, Institute of Radio Engineers, 1929.

Salzberg Bernard: See PROCEEDINGS for January, 1935.

Travis, Charles: Received B. S. and Ph. D. degrees, University of Pennsylvania. Engaged in nonradio activity until 1926. Engineer, Atwater Kent Manufacturing Company, 1926-1931; engineer, RCA License Division Laboratory, 1931-1935; Philco Radio and Television Corporation, 1935. Associate member, Institute of Radio Engineers, 1929.